

Durham E-Theses

Toxin-antitoxin system promoter binding by type IV antitoxins of Mycobacterium tuberculosis

BECK, IZAAK,NATHAN

How to cite:

BECK, IZAAK,NATHAN (2019) *Toxin-antitoxin system promoter binding by type IV antitoxins of Mycobacterium tuberculosis*, Durham theses, Durham University. Available at Durham E-Theses Online:
<http://etheses.dur.ac.uk/13076/>

Use policy



This work is licensed under a [Creative Commons Attribution Non-commercial 3.0 \(CC BY-NC\)](https://creativecommons.org/licenses/by-nc/3.0/)

Toxin-antitoxin system promoter binding by type IV antitoxins of *Mycobacterium tuberculosis*

A thesis submitted for the degree of MSc by Research

Izaak N. Beck
Biosciences
Grey College
Durham University



Grey College



Durham
University

Table of Contents

| | Page |
|--|-------------|
| Abstract | 5 |
| Declaration | 6 |
| Statement of Copyright | 6 |
| Acknowledgements | 7 |
| List of Figures | 9 |
| List of Tables | 10 |
| List of abbreviations | 11 |
| 1. Introduction | 13 |
| 1.1 Toxin-Antitoxin Systems | 14 |
| 1.1.1 Type I | 14 |
| 1.1.2 Type II | 18 |
| 1.1.3 Type III | 18 |
| 1.1.4 Type IV | 19 |
| 1.1.5 Type V | 20 |
| 1.2 Importance of <i>M. tuberculosis</i> type IV toxin-antitoxin systems | 22 |
| 1.3 DNA-binding in toxin-antitoxin systems | 23 |
| 1.3.1 DNA-binding and autoregulation in type II toxin-antitoxin systems | 24 |
| 1.3.1.1 DNA-binding in the RelBE system | 24 |
| 1.3.2 DNA-binding by type IV antitoxins | 25 |
| 2. Materials and methods | 27 |
| 2.1 Media, reagents and solutions | 27 |
| 2.2 Bacterial strains and culture | 27 |
| 2.3 Recombinant DNA techniques | 27 |
| 2.3.1 DNA purification and visualization | 27 |
| 2.3.1.1 Bacterial plasmid extraction | 27 |
| 2.3.1.2 | 27 |
| 2.3.2 Polymerase chain reaction (PCR) | 34 |
| 2.3.3 Cloning | 34 |

| Table of contents | Page |
|--|-------------|
| 2.3.3.1 Vector digest | 34 |
| 2.3.3.2 Insert amplification | 35 |
| 2.3.3.3 LIC reaction | 35 |
| 2.3.3.4 Annealing and transformation | 35 |
| 2.4 EMSA probe production | 37 |
| 2.5 Sequencing and sequence analysis | 37 |
| 2.6 Protein expression and purification | 42 |
| 2.6.1 Large scale protein expression | 42 |
| 2.7 Protein purification | 42 |
| 2.7.1 Isolating the soluble fraction | 42 |
| 2.7.2 Nickel-affinity chromatography, tag cleavage, and ortho-Nickel | 42 |
| 2.7.3 Fast protein liquid chromatography (FPLC) | 43 |
| 2.7.3.1 Size exclusion chromatography | 43 |
| 2.7.3.2 Anion exchange | 43 |
| 2.7.4 Protein storage | 43 |
| 2.8 Electrophoretic mobility shift assay (EMSA) | 45 |
| 2.8.1 EMSA visualisation and data processing | 45 |
| 2.9 Protein crystallisation | 46 |
| 2.9.1 Crystallisation screens | 46 |
| 2.9.2 Crystallisation screen optimization | 46 |
| 2.9.3 Crystal harvesting | 46 |
| 2.10 X-ray crystallography | 47 |
| 2.10.1 Data collection | 47 |
| 2.10.2 Data processing | 47 |
| 3. Results | 49 |
| 3.1 Antitoxin protein expression and purification | 49 |
| 3.2 DNA-binding via EMSA | 49 |
| 3.2.1 AbiEi binds to a single inverted repeat in the <i>AbiEi</i> promoter and displays positive cooperativity | 49 |
| 3.2.2 Rv1044 binding site is yet to be identified | 54 |

| Table of contents | Page |
|---|-------------|
| 3.2.3 Rv2827c binds to two inverted repeats in the <i>rv2827c</i> promoter | 54 |
| 3.2.3.1 Rv2827c binding to <i>rv2827c</i> -1 to -71 requires further study | 54 |
| 3.2.3.2 Rv2827c binds in a negatively cooperative manner to <i>rv2827c</i> -61 to -131 | 55 |
| 3.2.3.3 Rv2827c binds in a negatively cooperative manner to the -1 to -131 promoter region of <i>rv2827c</i> | 59 |
| 3.2.3.4 Negative cooperativity within the <i>rv2827c</i> -1 to -131 promoter region is DNA-sequence based rather than protein-protein interaction based | 59 |
| 3.2.4 Binding of type IV antitoxin proteins to non-cognate type IV toxin-antitoxin promoter regions | 64 |
| 3.2.4.1 Rv1044 can bind to the <i>abiEi</i> WT promoter region IRs | 64 |
| 3.2.4.2 Non-cognate promoter binding is unique to Rv1044- <i>abiEi</i> | 64 |
| 3.3 Crystallography | 67 |
| 3.3.1 Crystallisation | 67 |
| 3.3.2 X-ray crystallography data collection | 67 |
| 3.3.2.1 Structure solving | 67 |
| 4. Discussion | 70 |
| 4.1 Promoter region binding by type IV TA system antitoxins | 70 |
| 4.1.1 AbiEi-promoter interactions | 70 |
| 4.1.2 Mycobacterial antitoxin-promoter interactions | 71 |
| 4.1.2.1 Rv1044 promoter interactions | 71 |
| 4.1.2.2 Rv2827c promoter interactions | 72 |
| 4.2 Crystallography | 76 |
| 4.3 Future directions | 78 |
| Additional information, post the research period | 81 |
| Bibliography | 82 |

Abstract

Tuberculosis remains a global health concern that requires difficult and lengthy treatment regimens for latent and persistent infection. Toxin-antitoxin systems have been linked to controlling bacterial growth rate and also implicated in bacterial persistence. *Mycobacterium tuberculosis* carries three uncharacterised type IV toxin-antitoxin systems, *rv0837c/rv0836c*, *rv1044/rv1045*, and *rv2827c/rv2826c*, which are regulated during macrophage infection. This work characterises the DNA-binding capabilities of *Mycobacterium tuberculosis* protein antitoxins Rv0837c, Rv1044, Rv2827c, and a homologue AbiEi from *Streptococcus agalactiae*, with the aim of developing the autoregulatory paradigm for type IV antitoxins. Biochemical analysis of AbiEi corroborated already published work and structural characterisation of AbiEi was begun. Rv2827c demonstrated an ability to bind four sites within the *rv2827c* cognate promoter with differing affinities, and is the first example of a negatively cooperative autoregulatory response within toxin-antitoxin systems. Rv0837c could not be expressed in *E. coli* and Rv1044 could only bind to the *abiEi* promoter indicating a structural relationship to AbiEi. The functional relevance of these data are unclear, but given the essentiality of Rv2827c these results prompt further study into this family of systems.

Declaration

The work presented in this thesis was carried out at Durham University between September 2017 and September 2018. The work is my own original research unless otherwise stated or indicated by citation. This work has not been submitted for any other qualification.

Statement of copyright

The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior written consent and information derived from it should be acknowledged.

Acknowledgements

The work presented in this thesis was made possible by a number of people, to whom I am extremely grateful.

Firstly, I must thank my supervisor, Dr Timothy (stage name Tim) Blower for his excellent guidance and support throughout this project. Tim is always available for a chat, scientific or social, and this made the lab extremely easy to transition into. His attitude toward each of our projects is often overly optimistic and must be balanced out by a degree or two of pessimism, but his hunger for data pushes us all to get the work done. Tim is a supervisor that his students genuinely want to work for and I look forward to continuing this for the next four years as I progress to a PhD with him.

Thank you to my co-supervisor Dr Gary Sharples; he was always there when I needed him as a backup and gave invaluable insight into the EMSA I so long struggled to optimise. Turns out 2 hours at 200V will run everything off the end of a 7% gel. His cuddly bacteria/viruses are also a personal favourite and have been since my first year of undergraduate here. I hope there will never be another time in my life sometime shouts “YOU HAVE EBOLA” at me.

A large thank you to my thesis committee, Dr Peter Chivers and Dr Junli Liu for their guidance, support, and constructive criticism driving the project forward.

To Ben Usher and David Picton; thank you for creating such a welcoming and friendly environment to work in. I have the utmost respect for their dedication to maintaining a positive working environment in the lab where nobody is discriminated against. Their polar opposite working styles were a huge influence early on and I hope I found a middle ground between them. Thanks to both of you for all your help throughout the project and I look forward to working with you both for the next few years.

Thank you to Dr Ehmke Pohl for his guidance and short lectures on the dark arts of crystallography. In all honesty, most of this went over my head but one day I hope to

find my own way to understand it. Thank you to Dr Stefani Pohl, Charlie Tomlinson, and Rebecca Eno of the Pohl group for their help in crystallography data collection and processing; I genuinely would not have managed any of it without you all.

Finally, I must thank all the members of Lab 234 for making the lab such a great place to work, even in summer when we compete with the Biosciences greenhouse. I'm very glad to be staying on. Sorry if you don't feel that way too.

List of figures

| Figure | Page |
|--|-------|
| 1. Introduction | |
| 1.1 Toxin-antitoxin systems; types I – V in bacteria | 16-17 |
| 2. Materials and methods | |
| 2.1 Ligation independent cloning protocol | 36 |
| 2.2 Anion exchange chromatogram | 44 |
| 3. Results | |
| 3.1 Protein expression and purification | 51 |
| 3.2 DNA-binding studies of AbiEi with <i>abiEi</i> -1 to -71 | 52 |
| 3.3 Saturation curves and Hill plots for individual inverted repeats of the <i>abiEi</i> promoter -1 to -71 region | 53 |
| 3.4 DNA-binding studies of Rv1044 with <i>rv1044</i> -1 to -131 | 56 |
| 3.5 Consensus motif for <i>rv2827c</i> inverted repeats 1 – 4 | 56 |
| 3.6 DNA-binding studies of Rv2827c with <i>rv2827c</i> -1 to -71 | 57 |
| 3.7 DNA-binding studies of Rv2827c with <i>rv2827c</i> -60 to -131 | 58 |
| 3.8 DNA-binding studies of Rv2827c with <i>rv2827c</i> -1 to -131 | 60 |
| 3.9 DNA-binding studies of Rv2827c with <i>rv2827c</i> -1 to -71 IR4 perfect repeat | 61 |
| 3.10 Saturation curves and Hill plots for individual inverted repeats of the <i>rv2827c</i> promoter -1 to -131 region | 62-63 |
| 3.11 DNA-binding studies of Rv1044 with <i>abiEi</i> -1 to -71 | 65 |
| 3.12 DNA-binding studies of protein antitoxins with non-cognate promoter region probes | 66 |
| 3.13 AbiEi crystals and diffraction pattern | 68 |
| 4. Discussion | |
| 4.1 Schematic for <i>rv2827c</i> promoter assay with a dual reporter | 77 |

List of tables

| Table | Page |
|--|-------|
| 1. Introduction | |
| 1.1 Example toxin-antitoxin systems | 21 |
| 1.2 Essentiality of <i>M. tuberculosis</i> type IV toxin-antitoxin genes | 23 |
| 2. Materials and methods | |
| 2.1 Media used in this study | 28 |
| 2.2 Antibiotics and supplements used in this study | 29 |
| 2.3 Solutions used in this study | 30-32 |
| 2.4 Bacterial strains used in this study | 33 |
| 2.5 Primers used in this study | 38 |
| 2.6 EMSA probe oligos used in this study | 39-40 |
| 2.7 Plasmids used in this study | 41 |
| 2.8 Crystallisation screens for AbiEi | 48 |
| 3. Results | |
| 3.1 Data collection and refinement statistics | 69 |

List of abbreviations

| Abbreviation | Definition |
|------------------------|---|
| Å | Ångstrom |
| Ap | Ampicillin |
| bp | Base pair |
| BSA | Bovine serum albumin |
| CTD | C-terminal domain |
| DNA | Deoxyribonucleic acid |
| dNTP | Deoxyribonucleotide triphosphate |
| <i>E. coli</i> | <i>Escherichia coli</i> |
| EDTA | Ethylenediaminetetraacetic acid |
| EMSA | Electrophoretic mobility shift assay |
| g | Grams |
| HCl | Hydrogen chloride |
| IPTG | Isopropyl β – D – thiogalactopyranoside |
| IR | Inverted repeat |
| KCl | Potassium chloride |
| Kd | Dissociation constant |
| kDa | Kilodalton |
| Km | Kanamycin |
| LB | Luria broth |
| LIC | Ligation independent cloning |
| <i>M. tuberculosis</i> | <i>Mycobacterium tuberculosis</i> |
| mM | Millimolar |
| NA | Not available |
| NaCl | Sodium chloride |
| Ni-NTA | Nickel nitrilotriacetic acid |
| NTD | N-terminal domain |
| OD | Optical density |
| PAGE | Polyacrylamide gel electrophoresis |
| PCR | Polymerase chain reaction |
| PDB | Protein data bank |
| Poly(d[IC]) | Poly(deoxyinosinic-deoxycytidylic) acid |
| RNA | Ribonucleic acid |
| rpm | Revolutions per minute |
| <i>S. agalactiae</i> | <i>Streptococcus agalactiae</i> |
| SDS | Sodium dodecylsulphate |
| SEM | Standard error of mean |
| TA | Toxin-antitoxin |
| TB | Tuberculosis |
| TBE | Tris Boric acid EDTA |

List of abbreviations continued

| Abbreviation | Definition |
|--------------|-------------------------|
| Tc | Tetracycline |
| wHTH | Winged helix-turn-helix |
| WT | Wild type |
| μL | Microlitre |
| μm | Micrometre |
| μM | Micromolar |

Chapter 1. Introduction

Tuberculosis is a major public health concern with approximately 8.6 million new cases and 1.3 million deaths each year (Sulis et al., 2014). The causative agent, *Mycobacterium tuberculosis*, has evolved to become a successful human pathogen through the ability to evade host immune responses and establish a persistent infection with no clinical symptoms; as a result, the vast majority of cases exist as latent tuberculosis (van Crevel et al., 2002). The prevalence of latent infections and the requirement for a six-month course of antibiotics to treat infections is likely due to the existence of specialised “persister” cells (Wayne, 1994).

Persister cells are a subpopulation that have become phenotypically drug-tolerant due to the slowing or halting of cellular processes normally targeted by antibiotics. These cells account for about 1% of cells during stationary phase growth and in biofilms of most bacterial species (Lewis, 2008). The phenotypic change to a persister cell and subsequent reversion to an active cell are dynamic processes involving numerous and varied molecular pathways. Elucidating these pathways may offer new insights into approaches to tuberculosis treatment, effectively targeting active and persistent cells to eradicate the infection and reducing chronicity.

Specific toxin-antitoxin (TA) systems have been directly linked to the persister state (Holden and Errington, 2018; Wang and Wood, 2011; Wen et al., 2014) however, the global role of TA systems in persister formation is heavily debated (Goormaghtigh et al., 2018; Maisonneuve et al., 2018; Wang and Wood, 2011). TA systems are widespread and usually consist of a toxin that disrupts an essential cellular process, and an unstable and readily degraded antitoxin that inhibits toxicity (these systems will be discussed later). It is therefore possible to envisage how TA systems could integrate into molecular pathways governing persistence, by contributing to control of the cell’s metabolic rate.

It may therefore come as no surprise that *M. tuberculosis* boasts 88 TA systems (Sala et al., 2014a), with a subset of these systems becoming activated upon exposure to

stresses encountered during infection. This supports the theory that TA systems are important in bacterial adaptation to microenvironments (Ramage et al., 2009). Three functionally uncharacterised type IV TA systems are present in *M. tuberculosis*, Rv0836c/Rv0837c, Rv1044/Rv1045 and Rv2826c/Rv2827c (Sala et al., 2014b). These systems appear to be regulated in response to stresses such as those encountered during macrophage infection (Keren et al., 2011; Schnappinger et al., 2003; Torrey et al., 2016).

1.1 Toxin-Antitoxin Systems

All known toxin-antitoxin systems include a protein toxin and a labile RNA or protein antitoxin. The targets of the toxins are highly variable and the roles of TA systems in bacterial physiology are under increasing investigation (Goormaghtigh et al., 2018; Holden and Errington, 2018; Slayden et al., 2018). TA systems are characterised by the mode of toxin inhibition and are classed as types I - V (Figure 1). TA systems primarily exist as a bicistronic operon, with the downstream gene encoding a stable toxin and the upstream gene encoding a specific antitoxin in the form of an inherently unstable protein or RNA. A summary of known and studied TA systems can be found in Table 1.1.

1.1.1 Type I

The first discovered type of TA system exists as a convergently arranged overlapping gene pair with a cis-encoded sRNA antitoxin, or a divergently arranged gene pair with a trans-encoded antitoxin (Figure 1.1a) (Brantl, 2012). Toxicity is inhibited post-transcriptionally as the sRNA antitoxin binds in an antisense manner to the toxin transcript or a translationally coupled gene (such as *mok* as seen in the *hok/sok* system (Franch et al., 1997)) preventing toxin translation directly or indirectly, respectively (Thisted and Gerdes, 1992). Subsequently, the RNA duplex can be degraded or the ribosome binding site masked to further reduce toxin expression (Müller et al., 2016).

The toxins of Type I systems are generally accepted to be small hydrophobic proteins that interfere with membrane permeability due to a potential transmembrane domain; this ultimately impairs ATP synthesis by dissipating membrane potential.

Other type I toxins exist as RNase or DNase enzymes (Guo et al., 2014); the toxin SymE acts as an RNase rather than a membrane interfering protein and has been theorized to recycle damaged RNA in the SOS response or to prevent infection with RNA bacteriophages (Kawano et al., 2007), and Ra1R functions as a DNase in *E. coli* increasing resistance to fosfomycin (Guo et al., 2014). Many type I systems are located on prophages, which possibly explains their participation in protecting the cell from further phage infection, but also implicates them in maintenance of the prophage in the host chromosome (Durand et al., 2012).

Figure 1.1 Toxin-antitoxin systems; types I – V in bacteria

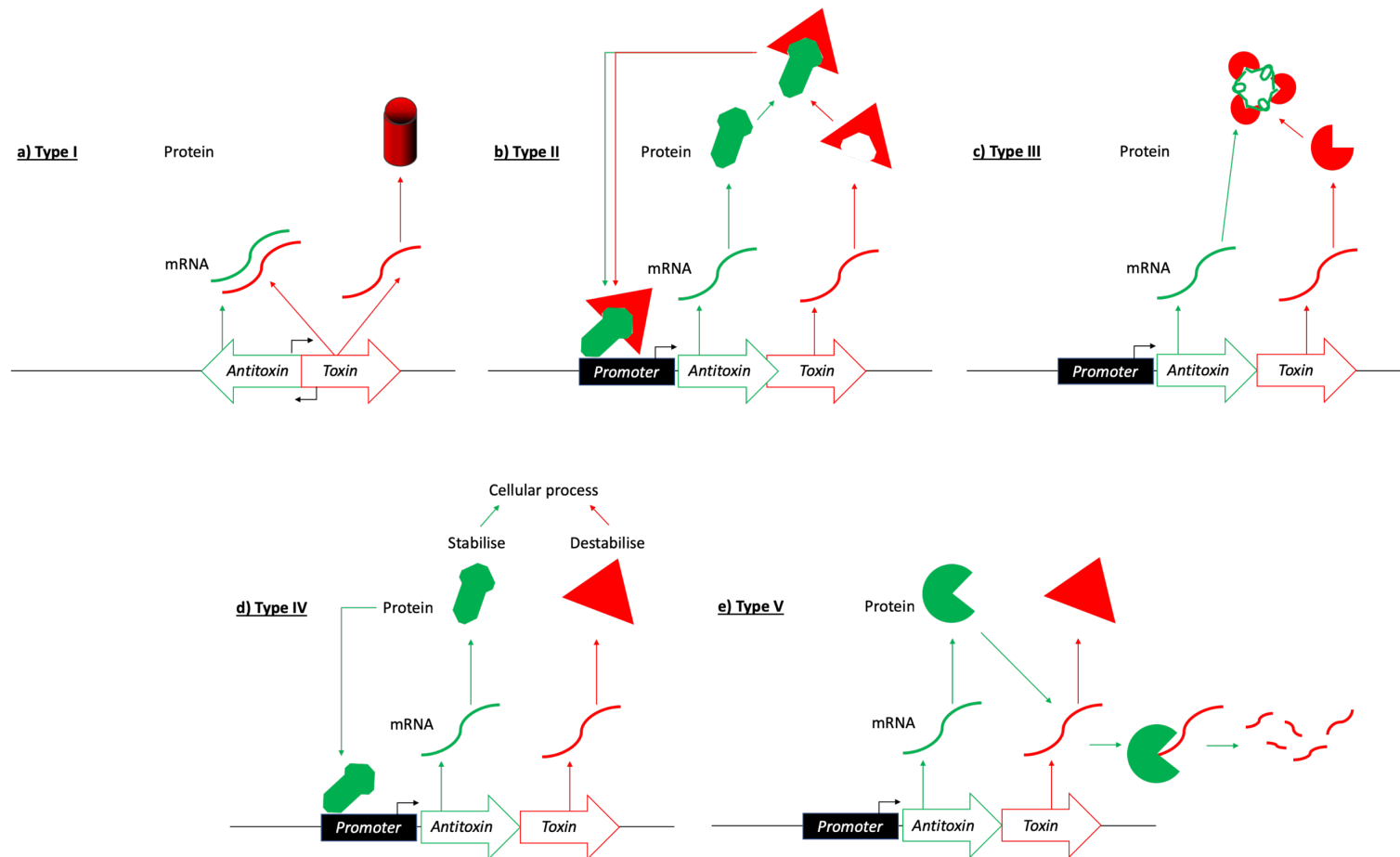


Figure 1.1 Legend overleaf

Figure 1.1 Toxin-antitoxin systems; type I – V in bacteria. a) Type I system antitoxicity occurs post-transcriptionally with the antitoxin sRNA binding toxin mRNA in an antisense manner, preventing translation. Type I toxins generally target the cell membrane; b) Type II system antitoxicity occurs post-translationally with the antitoxin protein binding directly to and inhibiting the toxin protein. Type II antitoxins are generally capable of negatively autoregulating the operon and type II toxins have a wide range of cellular targets inhibiting a variety of metabolic processes; c) Type III system antitoxicity occurs as a pseudoknotted antitoxin RNA binds to the protein toxin to inhibit toxicity; d) Type IV system antitoxicity occurs through an antagonistic mechanism whereby the antitoxin protein counteracts toxicity without directly interacting with the protein toxin. The emerging theory is that type IV antitoxins are also capable of autoregulation by binding to their cognate promoter regions; e) Type V system antitoxicity occurs via a protein antitoxin RNase targeting the toxin RNA, preventing toxin translation.

1.1.2 Type II

In Type II TA systems both the toxin and antitoxin are proteins that directly interact, which results in steric changes to the toxin or blocking of critical sites required for toxicity (Figure 1b). These systems have been structurally classified into six groups, namely Kid, Doc, VapC, RelE, HipA, and ζ (Blower et al., 2011a). The antitoxin proteins are intrinsically unstable and are usually comprised of a DNA-binding domain joined to a disordered regulatory region. This regulatory region is responsible for the instability and favoured proteolytic degradation of the antitoxin and becomes ordered when bound to the toxin counterpart (Buts et al., 2005; Cataudella et al., 2012; Garcia-Pino et al., 2010).

Typically, the genes form a bicistronic operon allowing for the majority of type II antitoxins to transcriptionally repress the toxin (alongside their own transcripts) via negative autoregulation. Activation of these systems depends on the rate of antitoxin degradation by Clp or Lon proteases and the outcome of transcriptional repression or de-repression is dictated by TA stoichiometry as demonstrated by the *RelBE* system; when the stoichiometry is such that RelB-RelE heterodimers form, or ternary RelB₂-RelE complexes, the locus is repressed. The heterotetramer, RelB₂-RelE₂, however, fails to bind DNA (Overgaard et al., 2008). Interestingly, a strictly pairwise system may not hold true for type II TA systems, rather a 'mix-and-match' principle may fit as described by Guglielmini & Van Melder (2011), after a bioinformatics approach revealed dramatic sequence diversity but structural conservation amongst toxins and prompted the inference that toxins may be capable of binding antitoxins of different classes. It has also been suggested that many families share a common toxin and regulatory unstructured antitoxin region, but different DNA-binding antitoxin domains have arisen through a series of illegitimate recombination events leading to TA system diversity (Leplae et al., 2011; Smith and Magnuson, 2004).

1.1.3 Type III

To date, only a few type III systems have been identified and validated with the toxin activity being inhibited by binding a non-coding pseudoknotted RNA (Figure 1c) (Goeders et al., 2016). ToxIN (*Pectobacterium atrosepticum*) (Fineran et al., 2009)

and AbiQ/AntiQ (*Lactococcus lactis*) (Emond et al., 1998) are systems with full length antitoxins ToxI and AntiQ containing 5.5 36-nucleotide and 2.8 35-nucleotide repeats respectively. These can be degraded to repeat monomers through sequence specific cleavage by the cognate toxin acting as an RNase. It has been shown that a single ToxI repeat is sufficient to repress ToxN but structural analysis revealed three pseudoknotted ToxI molecules bind to three ToxN protein toxins in a trimeric complex (Blower et al., 2011b). Furthermore, similar to many TA systems, negative autoregulation is observed for the ToxIN operon (Blower et al., 2009; Fineran et al., 2009; Samson et al., 2013). Toxins ToxN and AbiQ are homologues and intensive database searches identified a further two families; CptIN and TenpIN, in all of which “I” denotes the antitoxins and “N” denotes the toxin (Blower et al., 2012).

1.1.4 Type IV

In a somewhat outlying mechanism of inhibition to all other TA systems, type IV system components require no direct interaction for toxicity to be repressed (Figure 1d) (Masuda et al., 2012). In the first discovered system, the protein antitoxin CbeA and protein toxin CbtA of *E. coli* K-12 act antagonistically to stabilise and destabilise respectively cell shape and division proteins MreB and FtsZ (Tan et al., 2011). Interestingly, the stabilising activity of the antitoxin is capable of counteracting other MreB and FtsZ polymerisation inhibitors such as A22 for MreB and SulA and DicB for FtsZ (Masuda et al., 2012) demonstrating their general positive regulation capabilities. Various homologues of CbeA/CbtA exist in *E. coli* K-12; Ykfl/YafW and YpjF/YfjZ. It has been demonstrated that toxins Ykfl, YpjF, and CbtA bind FtsZ in *E. coli*. Interestingly, not only the cognate antitoxin rescues the cell from the appropriate toxin, but each of the homologues confer toxin resistance in a non-specific manner (Wen et al., 2017). This is likely due to the lack of toxin-antitoxin interaction and shared targets.

Additionally, it was shown that neither toxin nor antitoxin of the above type IV systems display autoregulation of their operon (Wen et al., 2017) which contrasts the type IV AbiEi/Eii system from *Streptococcus agalactiae* (Dy et al., 2014) in which two antitoxins (AbiEi) bind inverted repeats (IR) within the promoter via their N-terminal

domains to repress expression (Hampton et al., 2018). This type IV TA system is part of the abortive infection system family and is homologous to the previously described, uncharacterised putative type IV *M. tuberculosis* TA systems. Interestingly, the AbiEi C-terminal domain is sufficient for toxin neutralisation but also contributes to DNA-binding and transcriptional repression (Hampton et al., 2018).

1.1.5 Type V

A currently unique method of toxin inhibition is observed in the single defining TA module for type V systems (Figure 1e) (Wang et al., 2012). The *E. coli* GhoST system is comprised of toxin GhoT and sequence-specific endoribonuclease GhoS, primed for degradation of GhoT mRNA thereby offering post-transcriptional regulation of toxicity. GhoT is a small hydrophobic protein similar to toxins of the type I class and there is currently little evidence to support any autoregulation occurring (Wang et al., 2012).

Table 1.1 Example toxin-antitoxin systems

| System | Toxin | Antitoxin | Target/role | Reference |
|------------------------|--------|-----------|--|---------------------------------|
| <u>Type I</u> | | | | |
| <i>hok/sok</i> | Hok | Sok | Cell membrane/Plasmid stabilising | Thisted & Gerdes, 1992 |
| <i>symER</i> | SymE | SymR | mRNA cleaving (ribonuclease) | Kawano et al., 2007 |
| <u>Type II</u> | | | | |
| <i>ccdAB</i> | CcdB | CcdA | DNA gyrase/transcription arrest & dsDNA breaks | Jaffe et al., 1985 |
| <i>phd/doc</i> | Doc | Phd | 30S ribosome subunit/translation arrest | Jensen & Gerdes, 1995 |
| <i>mazEF</i> | MazE | MazF | mRNA cleaving (ribonuclease) | Hazan & Engleberg-Kulka, 2004 |
| <i>zeta/epsilon</i> | Zeta | Epsilon | Cell wall biosynthesis | Zielenkiewicz & Ceglowski, 2005 |
| <i>relBE</i> | RelE | RelB | mRNA degradation (ribonuclease) | Gotfredsen & Gerdes, 1998 |
| <i>vapBC</i> | VapC | VapB | RNA cleaving (ribonuclease) | Katz et al., 1992 |
| <i>higBA</i> | HigB | HigA | (ribonuclease) | Tian et al. 1996 |
| <u>Type III</u> | | | | |
| <i>toxIN</i> | ToxN | ToxI | Endoribonuclease | Fineran et al., 2009 |
| <u>Type IV</u> | | | | |
| <i>cbtA/cbeA</i> | CbtA | CbeA | FtsZ/MreB cytoskeletal protein | Masuda et al., 2012 |
| <i>ykfl/yafW</i> | Ykfl | YafW | FtsZ cytoskeletal protein | Brown & Shaw, 2003 |
| <i>abiEi/abiEii</i> | AbiEii | AbiEi | Unknown | Dy et al., 2014 |
| <u>Type V</u> | | | | |
| <i>ghoST</i> | GhoT | GhoS | Cell membrane | Wang et al., 2012 |

A summary of known, functionally characterised, toxin-antitoxin systems. These systems are present in a range of host organisms and have been split into the five categories defined by the mechanism of antitoxin action.

1.2 Importance of *M. tuberculosis* type IV toxin-antitoxin systems

The type IV TA systems present in *M. tuberculosis* have appeared in a few datasets analysing the transcriptome fluctuations during stresses encountered in macrophage infection and antibiotic pressure (Gupta et al., 2017). Further to this, the regulation of these genes has been analysed in persister cells and in response to *in vitro* stresses such as antibiotics (Keren et al., 2011; Sala et al., 2014b; Torrey et al., 2016). The inference is that large changes in a gene's expression suggest the gene must be important during this stage of the infection and in response to specific environmental conditions.

Considerable overlap of regulated genes identified by three separate studies reinforces the reliability of the studies and shows similarities in elicited responses; cells infecting the macrophage (Gupta et al., 2017), *in vitro* antibiotic pressure (Keren et al., 2011), and clinical persister isolates (Torrey et al., 2016) all result in the regulation of 10 genes by greater than 4-fold, one of which was Rv0837c. This highlights the potential importance of this gene during *M. tuberculosis* infection.

Conversely, data presented on the 10 most induced TA systems in drug-tolerant *M. tuberculosis* persister cells does not support the potential importance of *rv0837c*; rather, a combination of VapBC, RelBE, HigBA and other uncharacterised TA genes appeared to be upregulated (Sala et al., 2014). This potentially highlights the stochastic nature and variability of entry into the persister state. It is likely that a combination of pathways is required rather than a single gene and that there is more than one combination possible for entry into persistence.

The essentiality of the type IV TA systems was assayed by rendering them inactive through Himar1-based transposon mutagenesis assays (Griffin et al., 2011; Sassetti et al., 2003). Table 1.2 summarises the results and highlights antitoxin *rv2827c* as an essential gene for *M. tuberculosis* growth, potentially highlighting the bactericidal efficacy of the counterpart toxin product.

Table 1.2 Essentiality of *M. tuberculosis* type IV toxin-antitoxin genes

| | Sasseti <i>et al.</i> (2003) | Griffin <i>et al.</i> (2011) |
|----------------|-------------------------------------|-------------------------------------|
| Gene | Essential/Non-essential | Essential/Non-essential |
| <i>rv0836c</i> | Non-essential | Non-essential |
| <i>rv0837c</i> | Non-essential | Non-essential |
| <i>rv1044</i> | NA | Essential |
| <i>rv1045</i> | Non-essential | Non-essential |
| <i>rv2826c</i> | Non-essential | Non-essential |
| <i>rv2827c</i> | Essential | Essential |

NA – Not available

1.3 DNA-binding in toxin-antitoxin systems

Studying the regulation of toxin-antitoxin systems allows us to understand how expression of these toxic products is controlled by bacterial cells. A major feature of TA system control is autoregulation by protein antitoxins, which acts as a reinforcing mechanism to suppress counterpart toxicity and manage activation of the TA system.

Specific sequences within promoter regions, such as palindromic repeats, provide recognisable binding motifs for DNA-binding proteins (Murre *et al.*, 1989). These sequences offer a specific pattern of hydrogen bond donors and acceptors, on top of the natural helix distortion created by the DNA sequence from the region (Brennan and Matthews, 1989). For a TA system to be autoregulated, at least one of its components must be capable of binding to DNA, most likely somewhere in the cognate promoter region. It is well established that classical type II antitoxins act bi-functionally; first, to directly interact with the toxin counterpart as described before and secondly, to repress the operon by binding to an operator region within the system's promoter (Chan *et al.*, 2016).

DNA-binding antitoxin proteins in bacteria contain one of only a few recognised DNA-binding motifs within their DNA-binding domains, the most common motif is the ribbon helix-helix having been structurally identified in ParD (Oberer *et al.*, 2007), CcdA (Madl *et al.*, 2006) and RelB (Bøggild *et al.*, 2012). These proteins exist as dimers where two antiparallel β -strands form the ribbon, with each strand contributed by a

monomer. These strands are important for dimer formation and specific DNA base interactions (Milla et al., 1995; Raumann et al., 1994).

The helix-turn-helix motif is common in transcription factors and can be seen in the antitoxin HigA (Schureck et al., 2014). This motif is comprised of two alpha helices, one responsible for DNA-sequence recognition, the other responsible for stabilising the interaction (Brennan and Matthews, 1903). One sub-classification of the helix-turn helix is the winged helix-turn-helix (wHTH) which is formed by three alpha helices with the third helix being responsible for DNA-recognition (Brennan and Matthews, 1903). This motif has been identified in the uncharacterised putative type IV *M. tuberculosis* antitoxin Rv2827c (Janowski et al., 2009) and is believed to be present in the functionally characterised AbiEi antitoxin (Hampton et al., 2018).

1.3.1 DNA-binding and autoregulation in type II toxin-antitoxin systems

Within type II systems, autoregulation is classically controlled by a complex of the antitoxin and toxin binding to an operator site within the promoter (Afif et al., 2001). The efficacy of the antitoxin binding to its specific DNA sequence is modulated by the presence of toxin alongside the stoichiometry of the toxin and antitoxin (Garcia-Pino et al., 2010). As previously mentioned, toxin neutralisation by direct binding to the unstructured antitoxin region causes a switch to high affinity state for the operator site. Type II systems therefore offer a paradigm for protein antitoxin-DNA interactions in TA system repression.

1.3.1.1 DNA-binding in the RelBE system

For the type II RelBE system, when the stoichiometry is at RelB₂-RelE the operon is repressed, however, when high levels of RelE are achieved (in relation to antitoxin) the system becomes de-repressed in *E. coli*. (Overgaard et al., 2008). The importance of the toxin:antitoxin ratio in this system can be demonstrated via electrophoretic mobility shift assay (EMSA) using defined ratios and a promoter DNA-sequence probe (Overgaard et al., 2008). RelB alone shows a very weak affinity for the promoter

probe, however, on titrating in toxin to create protein complexes the affinity appeared to increase as seen by well-defined band shifts at low concentrations (Overgaard et al., 2008). Multiple binding sites appear in the *relBE* promoter and using a 2:1 molar ratio, positive cooperative binding is seen, as indicated by a sigmoidal binding curve (Overgaard et al., 2008). This creates a rapid switch from an active system to a repressed one. Inside this promoter the operator site (*relO*) was identified as an inverted repeat of 12 bp that is bound by the RHH motif of RelB₂ (Overgaard et al., 2009).

1.3.2 DNA-binding by type IV antitoxins

To date, AbiEi from *Streptococcus agalactiae* remains the only functionally characterized type IV system antitoxin proven capable of DNA-binding and negative autoregulation of its own operon. The C-terminal domain has a conserved charged surface that contributes to specific DNA-binding by the N-terminal domain to two inverted repeats in a positively cooperative manner, demonstrated by EMSAs and accompanying saturation curves and Hill plots (Hampton et al., 2018). Positive cooperative binding indicates that once one ligand (in this case AbiEi protein) binds to the target, the binding of subsequent ligands becomes easier (Crouch & Klee, 1980; Stefan & Le Novère, 2013). Cooperativity can be displayed using the Hill plot and calculated as the Hill coefficient (slope) whereby the maximum degree of cooperativity is equal to the number of binding sites with a coefficient of greater than 1 indicating positive cooperativity and less than 1 indicating negative cooperativity (Dahlquist, 1978; Hill and Av., 1910). It is predicted that AbiEi bends the DNA by 72° within the promoter by binding the operator site; this mechanism of repression was suggested to be common in the widespread AbiEi antitoxin family (abortive infection system members) (Hampton et al., 2018).

The AbiEi homologue, Rv2827c, is essential for *M. tuberculosis* growth *in vitro* and regulated in response to stress (Griffin et al., 2011; Keren et al., 2011; Sassetti et al., 2003; Torrey et al., 2016). The structure of Rv2827c has been solved to 1.93 Å and predicted to also be a DNA-binding protein (Janowski et al., 2009); the presence of a

wHTH in the NTD strongly supports this with the helix $\alpha 3$ sequence fragment 48-Pro-Asp-Ser-Ala-Ile-Arg-Glu-Leu-Arg-Arg-Ile-58 likely responsible for the sequence specific DNA-interaction. Initial DNA-binding tests further supported this (Janowski et al., 2009) however no sequences from the cognate promoter were tested therefore, we have no insight into the potential autoregulatory capacity of this antitoxin.

There are a number of gaps in our understanding regarding the four described AbiE family member antitoxins, AbiEi; Rv1044; Rv2827c; and Rv0837c. There are significant pieces of data for selected examples. For instance, studies of Rv2827c offer good insights and a model approach to the structural classification and analysis of these antitoxins (Janowski et al., 2009), and studies of AbiEi offer insights into the biochemistry and DNA-binding capabilities of this family (Dy et al., 2014; Hampton et al., 2018). It is therefore reasonable to pursue the characterisation of the remaining antitoxins to comprehensively cover these aspects for all target family members. Rv2827c, Rv1044 and Rv0837c all remain functionally uncharacterised in promoter binding. AbiEi remains structurally unclassified (alongside Rv1044 and Rv0837c). The aim of this body of work is therefore to add to our current understanding of the type IV protein antitoxins, through structural and functional studies of DNA-binding.

Type IV Antitoxin DNA-binding will be analysed *in vitro* using the electrophoretic mobility shift assay (EMSA). This has proven to be a useful technique in characterising the type IV antitoxin-operator site interactions and offers useful insight and support to the theory of autoregulation by type IV antitoxins via kinetics data (Hampton et al., 2018). It is hypothesised that the antitoxins will bind to operator sites within their own promoter regions, displaying similar kinetics and positive cooperativity as has been seen for AbiEi (Hampton et al., 2018). We also predict that, due to conserved domains, Rv2827c will serve as a useful model for solving the structure of AbiEi via molecular replacement.

Chapter 2. Materials and methods

2.1 Media, reagents and solutions

All media, antibiotics and other supplements, and solutions used in this study are detailed in Tables 2.1, 2.2 and 2.3, respectively. Where appropriate these were sterilised via autoclaving at 121 °C for 20 minutes or filtration via 0.22 µm filter.

2.2 Bacterial strains and culture

Both *E. coli* strains used in this study are listed in Table 2.4. *E. coli* were grown at 37 °C in liquid culture shaken at 300 rpm, or on agar plates. Culture temperature and shaking varied only in protein expression protocols as stated later. Growth was monitored using a cell density meter to (WPA Biowave C08000) give OD₆₀₀.

2.3 Recombinant DNA techniques

Molecular biology techniques involving DNA were performed by standard methods (Berger and Kimmel, 1987). All oligonucleotide primers and double-stranded oligonucleotide probe DNA sequences were obtained from Integrated DNA Technologies (IDT) and can be found in Tables 2.5 and 2.6 respectively.

2.3.1 DNA purification and visualization

2.3.1.1 Bacterial plasmid extraction

Plasmid DNA was purified using a NEB Monarch® Plasmid MiniPrep kit following the manufacturer's instructions. Plasmids were eluted in dH₂O for storage at -20 °C.

2.3.1.2 Agarose gel electrophoresis and DNA extraction

Agarose gel preparation is described in Table 2.3. A 6x DNA loading dye (New England Biolabs) was added to the DNA sample in the appropriate volume prior to loading into the agarose gel. The DNA molecules were separated by electrophoresis at 120 V until necessary resolution was achieved. Molecular weights were compared to a 1 kb ruler (ThermoFisher) and bands at the appropriate molecular weight cut out for DNA extraction using a NEB Monarch® Gel Extraction kit, following the manufacturer's instructions. DNA was eluted in dH₂O for storage at -20 °C.

Table 2.1 Media used in this study

| Medium | Ingredients per litre |
|------------------------------------|--|
| Luria-broth | 10 g Casein digest peptone 5 g Yeast extract 5 g NaCl |
| Luria-broth agar | 10 g Casein digest peptone 5 g Yeast extract 5 g NaCl 15 g Agar |
| 2 x YT broth (nutrient rich broth) | 16 g Casein digest peptone 10 g Yeast extract 5 g NaCl |

Table 2.2 Antibiotics and supplements used in this study

| Chemical (Abbreviation) | Stock solution | Working concentration |
|--|--|-----------------------|
| <hr/> | | |
| Antibiotic | | |
| Ampicillin (Ap) | 1000 x stock, 100 mg/mL in dH ₂ O, stored at -20 °C | 100 µg/mL |
| Supplement | | |
| Isopropyl β – D – thiogalactopyranoside (IPTG) | 1000 x stock, 1 M in dH ₂ O, stored at -20 °C | 1 mM |
| <hr/> | | |

Table 2.3 Solutions used in this study

| Solution | Components |
|--|--|
| DNA work | |
| 50x TAE Buffer (per L) | 242 g Tris base [tris(hydroxymethyl)aminomethane] 57.1 mL Glacial acetic acid (17.4 M) 18.61 g EDTA, disodium salt pH 8.0 |
| 6x DNA loading dye | |
| Agarose gel mix | 1% Agarose in 1 x TAE 500 ng Ethidium bromide per mL gel |
| 1D SDS-PAGE | |
| 10 x Stock electrode buffer (per L) | 30.2 g Tris base [tris(hydroxymethyl)aminomethane] 141 g Glycine 0.1 L 10% SDS pH 8.3 |
| Working electrode buffer | 100 mL Stock electrode buffer Diluted to 1000 mL |
| 15% (resolving) acrylamide gel (enough for four 10 cm BioRad gels) | 9 mL 40% Acrylamide 12 mL 0.75 M Tris pH 8.8 240 µL 10% SDS 240 µL 10% Ammonium persulphate 24 µL TEMED 2.5 mL dH ₂ O |
| 6% (stacking) acrylamide gel (enough for four 10 cm BioRad gels) | 1.5 mL 40% Acrylamide 1.5 mL 1.25M Tris-HCl pH 6.8 150 µL 10% SDS 150 µL 10% Ammonium persulphate 15 µL TEMED 11.7 mL dH ₂ O |

Table 2.3 continued. Solutions used in this study

| Solution | Components |
|--|---|
| Protein purification and crystallisation | |
| Lysis buffer (A500) | 500 mM NaCl 20 mM Tris HCl pH 7.9 5 mM Imidazole pH 8.0 10% Glycerol |
| Ni-NTA column elution buffer (B500) | 500 mM NaCl 20 mM Tris HCl pH 7.9 250 mM Imidazole pH 8.0 10% Glycerol |
| Ni-NTA column low salt elution buffer (B100) | 100 mM NaCl 20 mM Tris HCl pH 7.9 250 mM Imidazole pH 8.0 10% Glycerol |
| FPLC low salt buffer (A100) | 100 mM NaCl 20 mM Tris HCl pH 7.9 5 mM Imidazole pH 8.0 10% Glycerol |
| FPLC high salt buffer (C1000) | 1000 mM NaCl 20 mM Tris HCl pH 7.9 10% Glycerol |
| FPLC sizing column buffer | 500 mM KCl 50 mM Tris HCl pH 7.9 10% Glycerol |
| Protein sample storage buffer | 500 mM KCl 50 mM Tris HCl pH 7.9 70% Glycerol |
| Crystallisation buffer | 200 mM NaCl 20 mM Tris HCl pH 7.9 2.5 mM DTT |

Table 2.3 continued. Solutions used in this study

| Solution | Components |
|--|--|
| DNA-binding studies | |
| 5 x TBE (per L) | 54 g Tris base 27.5 g Boric Acid 20 mL EDTA 0.5 M pH 8.3 |
| 0.5 x TBE electrode buffer | 500 mM NaCl 20 mM Tris HCl pH 7.9 250 mM Imidazole pH 8.0 10% Glycerol |
| 5 x EMSA binding buffer | 750 mM KCl 50 mM Tris HCl pH 8.0 2.5 mM EDTA pH 8.0 0.5% Triton-X 100 1mM DTT 55% glycerol |
| 7 % resolving acrylamide gel (enough for four 10 cm BioRad gels) | 4.2 mL 40% Acrylamide 2.4 mL 5 x TBE 168 µL 10% ammonium persulphate 16.8 µL TEMED 17.2 mL dH ₂ O |
| 5 % resolving acrylamide gel (enough for four 10 cm BioRad gels) | 3 mL 40% Acrylamide 2.4 mL 5 x TBE 168 µL 10% ammonium persulphate 16.8 µL TEMED 18.4 mL dH ₂ O |

Table 2.4 Bacterial strains used in this study

| Strain | Genotype | Source |
|-------------------------|---|---------------------|
| <i>Escherichia coli</i> | | |
| DH5 α | <i>F- Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rk-, mk+) phoA supE44 λ-thi-1 gyrA96 relA1</i> | Invitrogen |
| ER2566 | <i>fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 Δ(mcrC-mrr)114::IS10</i> | New England Biolabs |

2.3.2 Polymerase chain reaction (PCR)

Q5 DNA polymerase was used for DNA amplification via PCR according to the components and steps outlined below. Each component can be scaled to alter the total reaction volume and the annealing temperature was dependent on specific primer pairs.

Q5 DNA polymerase PCR

| Component | Volume (μ L) | Step | Temp ($^{\circ}$ C) | Time |
|-----------------------|-------------------|-------------------------|----------------------|----------|
| 10 x Q5 Buffer | 5 | 1) Initial denaturation | 95 | 30 sec |
| 10 mM dNTP's | 1 | 2) Denaturation | 95 | 10 sec |
| 10 μ M Primer 1 | 2.5 | 3) Annealing | Varies* | 30 sec |
| 10 μ M Primer 2 | 2.5 | 4) Extension | 72 | 30 sec |
| DNA sample (template) | 1 | Repeat 2 – 4 x 40 times | | |
| Q5 polymerase | 0.5 | 5) Final extension | 72 | 2 min |
| dH ₂ O | 37.5 | 6) Hold | 10 | ∞ |
| | 50 | | | |

2.3.3 Cloning

Ligation independent cloning (LIC) (Aslanidis and de Jong, 1990) was performed in order to generate expression constructs from which the genes of interest can be selectively expressed. Antitoxin gene *abiEi* was cloned into pSAT1-LIC, fused to a 6-HIS-human SUMO-2 gene under control of the T7 promoter. The LIC protocol is summarised in Figure 2.1.

2.3.3.1 Vector digest

1 μ g of pSAT1-LIC was digested with *Stu*I producing a linearised plasmid with exposed blunt ended LIC sites. The linearised DNA was run on agarose gel and extracted as detailed in 2.3.1.2. Below is a summary of the digestion reaction.

pSAT1-LIC digestion with *Stu*I

| Component | Volume (μ L) | Temperature | Time |
|-----------------------|----------------------|-----------------|------|
| pSAT1-LIC | Variable (1 μ g) | | |
| 10 x Buffer 2.1 (NEB) | 2.5 | 37 $^{\circ}$ C | 2 hr |
| <i>Stu</i> I (NEB) | 1 | | |
| dH ₂ O | to 25 μ L | | |

2.3.3.2 Insert amplification

A LIC site flanked-*abiEi* was amplified via the Q5 polymerase chain reaction (PCR) using the method and components described previously with template plasmid pRLD30 containing *abiEi*. Primers were designed using the *abiEi* sequence with the addition of sequences complementary to the LIC sites of pSAT1-LIC and can be found as “FWD *abiEi* LIC” (TRB1048) and “REV *abiEi* LIC” (TRB1049) in table 2.5. Amplified inserts were then run on an agarose gel and purified as detailed in 2.3.2.1.

2.3.3.3 LIC reaction

Purified linearised vector and amplified insert, both with LIC sites, were carried forward to the LIC reaction to expose complementary ‘sticky ends’ on both vector and insert. The reaction for each differs slightly and is detailed below.

LIC reaction

| Vector | | Insert | | |
|-----------------------|-------------|-----------------------|-------------|-----------------|
| Component | Volume (μL) | Component | Volume (μL) | |
| Vector | 25 | Insert | 10 | |
| dTTP (25 mM) | 5 | dATP (25 mM) | 2 | |
| 10 x 2.1 Buffer (NEB) | 5 | 10 x 2.1 Buffer (NEB) | 2 | 1) 22 °C 30 min |
| DTT (100 mM) | 2.5 | DTT (100 mM) | 1 | 2) 75 °C 20 min |
| T4 DNA | 1 | T4 DNA | 0.4 | |
| Polymerase (NEB) | | Polymerase (NEB) | | |
| dH ₂ O | 11.5 | dH ₂ O | 4.6 | |
| | 50 | | 20 | |

2.3.3.4 Annealing and transformation

The products of the LIC reaction were combined in a ratio of 1:1 as 40% of the reaction volume. Annealing was carried out at room temperature overnight. The total annealing reaction volume was added to a 200 μL aliquot of heat-shock competent DH5α cells for transformation and selection on ampicillin supplemented LB agar. The recombinant plasmid was then harvested from cells using the NEB Monarch® MiniPrep kit.

Figure 2.1 Ligation independent cloning protocol

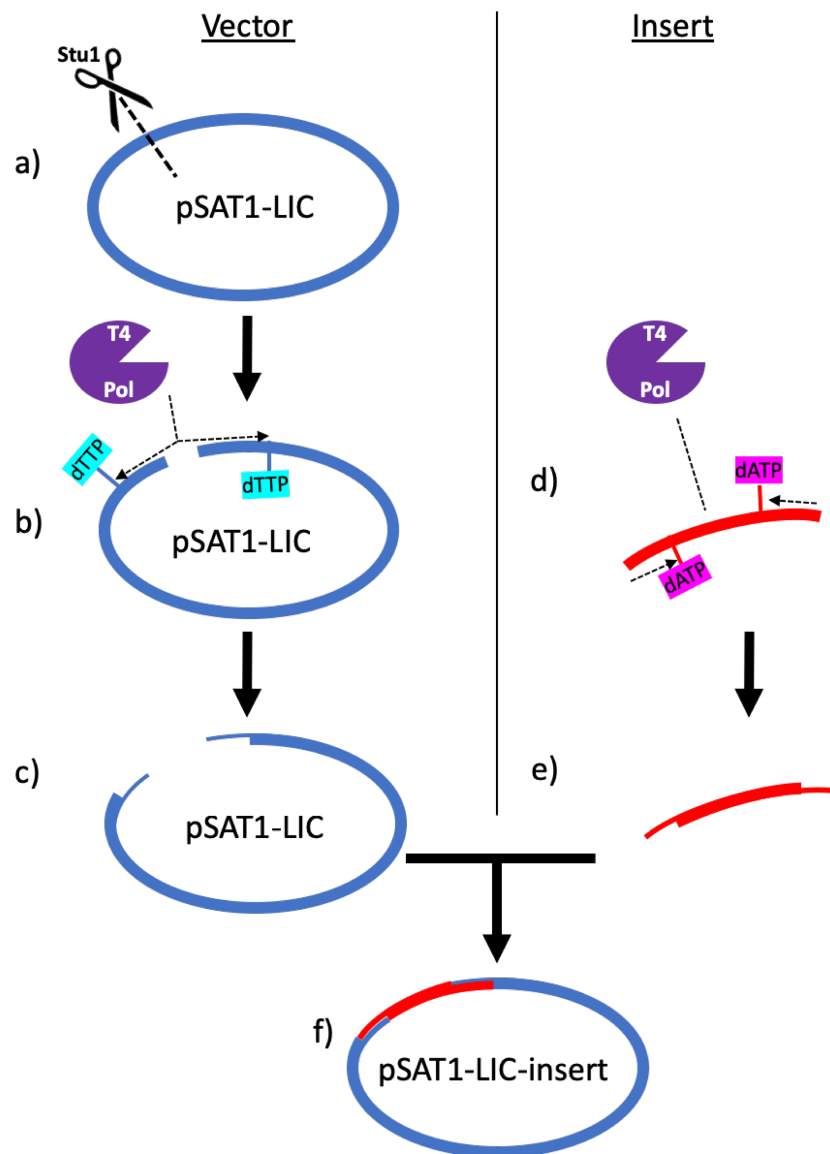


Figure 2.1 The ligation-independent cloning protocol with pSAT1-LIC. a) Purified pSAT1-LIC is digested using the restriction enzyme *StuI* at the AGGCCT palindromic site to create a linear, blunt-ended plasmid. b) the T4 DNA polymerase 3' – 5' exonuclease is used to create pSAT1-LIC with the long single strand overhang; c) pSAT1-LIC with exposed overhangs; d) insert DNA (gene of interest – GOI) is treated with T4 DNA polymerase, creating complementary overhangs to pSAT1-LIC; e) gene of interest with overhangs; f) pSAT1-LIC and the complementary gene of interest are mixed together, allowing the complementary overhangs to anneal and re-seal the vector with the GOI inserted.

2.4 EMSA probe production

Fluorescently labelled probes, alongside unlabelled competitor probes, were amplified by Q5 PCR. Primers for the reactions are listed in Table 2.5 and oligonucleotide primers (oligos) containing the probe sequences taken from gene promoters are listed in Table 2.6. A constant region was added to each oligo to which the reverse (REV) primer would anneal, creating either tagged or untagged double-stranded probe sequences. Probes were eluted in dH₂O and stored at -4 °C. The QS1 probe containing 131 base pairs upstream of *rv2827c* (with four identified binding sites) was created by Q5 PCR using forward primer TRB1107 and reverse primers TRB1067/TRB1068 as appropriate. The template DNA was plasmid pTRB484 containing a 500 base pair upstream region of *rv2827c* fused to *lacZ* (created for promoter assays – not in this study).

2.5 Sequencing and sequence analysis

Extracted plasmids containing the gene of interest were sequenced in-house by DBS Genomics, Durham University Biosciences Department, using the ABI 3730 DNA sequencer via primer walking (primers listed in Table 2.5). Confirmation of gene sequences was performed using 4Peaks and BLASTN sequence alignment (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Table 2.5 Primers used in this study

| Primer | Sequence | Notes (Organism/Gene) |
|---------|---|--|
| TRB873 | TTAATGCAGCTGATTAATACG | FWD pSAT LIC sequencing |
| TRB875 | TACTCAAGCTTATGCATGC | REV pSAT LIC sequencing |
| TRB1048 | CAACAGCAGACGGGAGGTTCAAAAAAGA GATTCTACTCGATTTTATAG | FWD <i>abiEi</i> LIC, <i>S. agalactiae</i> , <i>abiEi</i> |
| TRB1049 | GCGAGAACCAAGGAAAGGTTATTATATTA GAACCTCCAGAGTTTGTTTAAC | REV <i>abiEi</i> LIC, <i>S. agalactiae</i> , <i>abiEi</i> |
| TRB1065 | AAAAGAAAATGTTGCTTTTATACCACA | FWD for TRB1061, TRB1063, <i>S. agalactiae</i> , <i>abiEi</i> |
| TRB1066 | AAAAGAAAACCCCCCCCC | FWD for TRB1062, TRB1064, <i>S. agalactiae</i> , <i>abiEi</i> |
| TRB1067 | TGCGCACTGACAAAAGCTT | REV EMSA untagged |
| TRB1068 | /56-FAM/TGCGCACTGACAAAAGCTT | REV EMSA 56-FAM (fluorescein) tagged |
| TRB1087 | AACTAGGCGCGCCTAG | FWD for TRB1086, <i>M.</i> <i>tuberculosis</i> , <i>rv1044</i> |
| TRB1103 | GTATCTGCGACAAGGGCAG | FWD for TRB1102, <i>M.</i> <i>tuberculosis</i> , <i>rv1044</i> |
| TRB1105 | CAAGTGATTCTTGAGTTTGAACATTG | FWD for TRB1104, TRB1271, <i>M. tuberculosis</i> , <i>rv2827c</i> |
| TRB1107 | CAGGGCACTTGAGTTTGAAC | FWD for TRB1106, TRB1277, <i>M. tuberculosis</i> , <i>rv2827c</i> |
| TRB1272 | CAAGTGATTCCCCCCCC | FWD for TRB1271, TRB1274, <i>M. tuberculosis</i> , <i>rv2827c</i> |
| TRB1276 | CAGGGCCCCCCCC | FWD for TRB1275, TRB1278, <i>M. tuberculosis</i> , <i>rv2827c</i> |
| TRB1280 | CAAGTTCAATACCATCATAAAAAAGAAGG | FWD for TRB1279, <i>S.</i> <i>agalactiae</i> , <i>abiEi</i> |
| TRB1297 | AAGTGATTACTTGAATTCACACCG | FWD for TRB1296, <i>M.</i> <i>tuberculosis</i> , <i>rv2827c</i> |

Table 2.6 EMSA probe oligos used in this study

| Oligo | Sequence ^a | Notes (Organism / Gene) |
|---------|---|---|
| TRB1061 | AAAAGAAAATGTTGCTTTTATACCACAAAT ATTGTAAAATTGTAGTGTAAGCAACAAG TGGGGGGCCGTAAAGCTTTTGTCAAGTGCAGCA | <i>Streptococcus agalactiae</i> / <i>abiEi</i> -1 to -71 Wild type |
| TRB1062 | AAAAGAAAACCCCCCCCCCTACCACAAAT ATTGTAAAATTGTAGTGTAAGCAACAAG TGGGGGGCCGTAAAGCTTTTGTCAAGTGCAGCA | <i>Streptococcus agalactiae</i> / <i>abiEi</i> -1 to -71 Mutant; inverted repeat 1 poly-C track substitution |
| TRB1063 | AAAAGAAAATGTTGCTTTTATACCACAAAT ATTGTAAAATTGTAGTGCCCCCCCCCAGT GGGGGGCCGTAAAGCTTTTGTCAAGTGCAGCA | <i>Streptococcus agalactiae</i> / <i>abiEi</i> -1 to -71 Mutant; inverted repeat 1 & 2 poly-C track substitution |
| TRB1064 | AAAAGAAAACCCCCCCCCCTACCACAAAT ATTGTAAAATTGTAGTGCCCCCCCCCAGT GGGGGGCCGTAAAGCTTTTGTCAAGTGCAGCA | <i>Streptococcus agalactiae</i> / <i>abiEi</i> -1 to -71 Mutant; inverted repeat 1 & 2 poly-C track substitution |
| TRB1086 | AACTAGGCGCGCCTAGCCTGGACGAGTCCC CGGGCCGACATTCGCCCGAGGCCTTGGCCT CCATCACCTAAAAGCTTTTGTCAAGTGCAGCA | <i>Mycobacterium tuberculosis</i> H37Rv / <i>rv1044</i> , -1 to -71 Wild type |
| TRB1102 | GTATCTGCGACAAGGGCAGCGTCGATGCCT CGACATGCAGAGTCGGTGTTCGCTTACGC GAAGTAGGCGCAAGCTTTTGTCAAGTGCAGCA | <i>Mycobacterium tuberculosis</i> H37Rv / <i>rv1044</i> , -61 to -131 Wild type |
| TRB1104 | CAAGTGATTCTTGAGTTGAACATTGTTGC GTACAGATATAGTATAGTTCCGGTGTGAA TTCAAGTTCGAAGCTTTTGTCAAGTGCAGCA | <i>Mycobacterium tuberculosis</i> H37Rv / <i>rv2827c</i> , -1 to -71 Wild type |
| TRB1106 | CAGGGCACTTGAGTTTGGAAACGGGTTTCGT ACTGTCACTGACCGAAGCCCGTTCTAAAT CAAGTGATTTCGAAGCTTTTGTCAAGTGCAGCA | <i>Mycobacterium tuberculosis</i> H37Rv / <i>rv2827c</i> , -61 to - 131 Wild type |
| TRB1262 | CGGTCGTCCGCGAAGCGCGTTTTGTCTCTC GCCGTTGCACCGCATCGCCGCCAGCCCGTA CAGGGCACTTAAGCTTTTGTCAAGTGCAGCA | <i>Mycobacterium tuberculosis</i> H37Rv / <i>rv2827c</i> , -121 to - 191 Wild type |
| TRB1271 | CAAGTGATTCCCCCCCCCCCCCCCCCCCC CTACAGATATAGTATAGTTCCGGTGTGAA TTCAAGTTCGAAGCTTTTGTCAAGTGCAGCA | <i>Mycobacterium tuberculosis</i> H37Rv / <i>rv2827c</i> , -1 to -71 Mutant; inverted repeat 3 poly-C track substitution |
| TRB1273 | CAAGTGATTCTTGAGTTTGAACATTGTTGC GTACAGATATAGTACCCCCCCCCCCCCCCC CCCCCTCGAAGCTTTTGTCAAGTGCAGCA | <i>Mycobacterium tuberculosis</i> H37Rv / <i>rv2827c</i> , -1 to -71 Mutant; inverted repeat 4 poly-C track substitution |

^a Probe sequences are fused with a constant region from the lacZ gene, highlighted in grey. The reverse primer was designed to anneal to this sequence for amplification.

Table 2.6 continued. EMSA probe oligos used in this study

| Oligo | Sequence ^a | Notes (Organism / Gene) |
|---------|---|---|
| TRB1274 | CAAGTGATTCCCCCCCCCCCCCCCCCCCCC CTACAGATATAGTACCCCCCCCCCCCCCCCCC CCCCCTCGAAGCTTTTGTCAAGTGC | <i>Mycobacterium tuberculosis</i> H37Rv / rv2827c, -1 to -71 Mutant; inverted repeat 3 & 4 poly-C track substitution |
| TRB1275 | CAGGGCCCCCCCCCCCCCCCCCCCCCTA CTGTCACTGACCGAAGCCGTTCTAAATC AAGTGATTAAAGCTTTTGTCAAGTGC | <i>Mycobacterium tuberculosis</i> H37Rv / rv2827c, -61 to - 131 Mutant; inverted repeat 1 poly-C track substitution |
| TRB1277 | CAGGGCACTTGAGTTTGAACGGGTTTCGT ACTGTCACTGACCCCCCCCCCCCCCCCCC CCCCGATTAAAGCTTTTGTCAAGTGC | <i>Mycobacterium tuberculosis</i> H37Rv / rv2827c, -61 to - 131 Mutant; inverted repeat 2 poly-C track substitution |
| TRB1278 | CAGGGCCCCCCCCCCCCCCCCCCCCCTA CTGTCACTGACCCCCCCCCCCCCCCCCC CCCGATTAAAGCTTTTGTCAAGTGC | <i>Mycobacterium tuberculosis</i> H37Rv / rv2827c, -61 to - 131 Mutant; inverted repeat 1 & 2 poly-C track substitution |
| TRB1279 | CAAGTTCAATACCATCATAAAAAAGAAGGC TTGATTTTGTTCACATAAATTGAGGGA AAAGAAAATGTAAGCTTTTGTCAAGTGC | <i>Streptococcus agalactiae</i> / abiEi -61 to -131 Wild type |
| TRB1296 | AAGTGATTACTTGAATTCACACCGGAACT ATACAGATATAGTATAGTTTCCGGTGTGAA TTCAAGTTCGAAGCTTTTGTCAAGTGC | <i>Mycobacterium tuberculosis</i> H37Rv / rv2827c, -1 to -71 Mutant; inverted repeat 4 perfect inverted repeat |

^a Probe sequences are fused with a constant region from the lacZ gene, highlighted in grey. The reverse primer was designed to anneal to this sequence for amplification.

Table 2.7 Plasmids used in this study

| Plasmid | Backbone plasmid | Resistance | Inserted gene / region | Origin organism / notes | Source |
|---------|------------------|------------|-------------------------|--|------------------------------|
| pRLD30 | pTRB30 | Km | <i>abiEi</i> | <i>S. agalactiae</i> , antitoxin | Dy et al. 2014 (Fineran Lab) |
| pTRB525 | pSAT1-LIC | Ap | <i>abiEi</i> | <i>S. agalactiae</i> , antitoxin | This study |
| pTRB491 | pSAT1-LIC | Ap | <i>rv1044</i> | <i>M. tuberculosis</i> , antitoxin | Blower Lab |
| pTRB493 | pSAT1-LIC | Ap | <i>rv2827c</i> | <i>M. tuberculosis</i> , antitoxin | Blower Lab |
| pTRB484 | pRW50 | Tc | <i>rv2827c promoter</i> | <i>M. tuberculosis</i> , 500 bp of promoter fused to <i>lacZ</i> | Blower Lab |

2.6 Protein expression and purification

Using pSAT1-LIC as our expression vector allows for high scale protein expression exploiting the T7 polymerase system. Selecting a host strain such as ER2566 (Table 2.4), with a chromosomally integrated T7 RNA polymerase under the control of the *lac* promoter, allows for selective expression of our gene of interest under the control of the T7 promoter.

2.6.1 Large scale protein expression

The pSAT1-LIC derived expression vector (plasmids can be found in Table 2.7) was transformed into ER2566 and grown using pre-defined conditions in nutrient rich 2 x YT media supplemented with Ap to an optical density (OD_{600}) of 0.6 – 0.8. The cultures were cooled to 17 °C and IPTG was added to a final concentration of 1 mM to induce expression of the target protein. Cultures were shaken at 150 rpm for 16 hours to maximise expression.

2.7 Protein purification

Proteins were expressed and purified following published protocols (Blower et al., 2016), with small variations as appropriate.

2.7.1 Isolating the soluble fraction

Bacterial cells were pelleted from liquid culture by centrifugation at 4200 x g for 30 minutes at 4 °C. Cell pellets were resuspended in lysis buffer A500 and sonicated for a total of 3 min. The sonicated sample was centrifuged at 20,000 x g for 40 min at 4 °C to isolate the soluble fraction from cell debris.

2.7.2 Nickel-affinity chromatography, tag cleavage, and ortho-Nickel

The protein rich isolated soluble fraction was passed through a Ni-NTA His-Trap™ HP 5 mL column (GE Healthcare) at slow speed to maximise recombinant protein binding via the 6-His tag. A wash step was performed using lysis buffer. The column was eluted from using the high imidazole elution buffer B500. The eluted sample was dialysed for 16 hours into lysis buffer A500 with the addition of 40 mg SENP2 in order to cleave the 6His-hSUMO tag, therefore reducing the metal binding affinity of the protein. The SENP2 treated sample was then run through another Ni-NTA His-Trap™

HP 5 mL column (ortho-Ni step) and the flow through collected. The flow through sample now contains our cleaved, partially purified protein of interest.

2.7.3 Fast protein liquid chromatography (FPLC)

FPLC steps are carried out using an Äkta protein chromatography system (GE Healthcare). This allows for programmable, precise and semi-quantifiable fractionation with multiple buffers when required.

2.7.3.1 Size exclusion chromatography

A HiPrep column loaded with sephacryl S-300 HR SEC resin (GE Healthcare) was used to partially separate proteins by size. The column was equilibrated in sizing column buffer and a concentrated protein sample applied and fractionated at a rate of 0.5 mL/min. Column calibration allows us to identify proteins of specified and appropriate weights which can be visualised on an output chromatogram. Peaks on the chromatogram were sampled and analysed by SDS-PAGE. Fractions containing the protein of interest were carried forward for further purification if needed.

2.7.3.2 Anion exchange

The protein sample was loaded on to a HiTrap Q HP anion exchange 5 mL column (GE Healthcare) in low salt buffer A100. This column was then subjected to an increasing salt gradient using the Äkta system, titrating in high salt buffer C1000 until a final salt concentration of 500 mM NaCl was achieved. Fractions were collected and analysed by SDS-PAGE. Fractions containing the protein of interest were pooled. An example anion exchange fractionation chromatogram can be seen in Figure 2.2.

2.7.4 Protein storage

Purified protein in either FPLC sizing column buffer, or an appropriate salt concentration from anion exchange, was concentrated to 1.5 mg/mL. The sample was added to protein storage buffer in the ratio of 2:1 to reach a final protein concentration of 1 mg/mL and a final glycerol (cryoprotectant) concentration of 30% (v/v). The protein sample was snap-frozen in liquid nitrogen prior to storage at -80 °C.

Figure 2.2 Anion exchange chromatogram

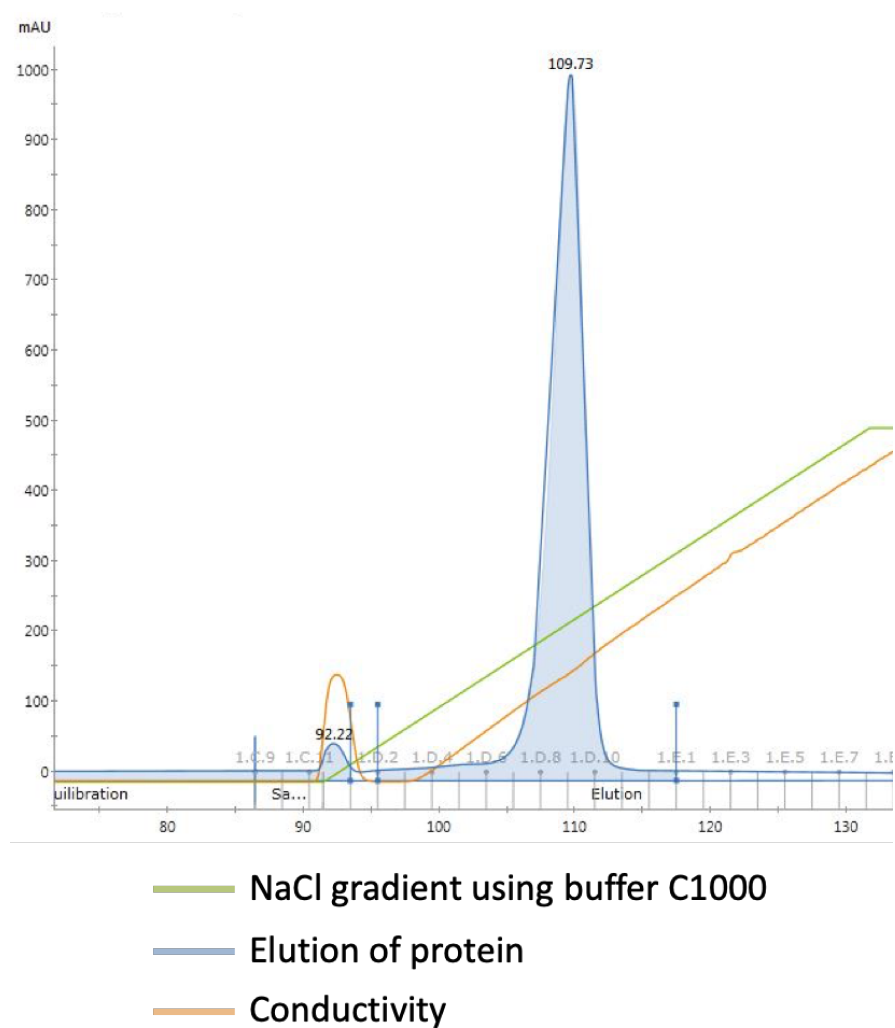


Figure 2.2 Anion exchange chromatogram from protein purification using the Äkta. The purified sample of protein elutes in a tight peak in response to a specific salt concentration. Protein is measured by UV and registered in milli-arbitrary units. Fractions are collected in a 96 well block with fractions corresponding to the X-axis.

2.8 Electrophoretic mobility shift assay (EMSA)

Promoter region sequence containing probes were amplified as previously described. Proteins were diluted to appropriate concentrations using diluent buffers matching their storage buffer constitution. Each individual binding reaction included 250 fmoles of fluorescently labelled probe. The reaction components are shown below. Native polyacrylamide gels (at either 7% or 5% as required) were pre-run at 150 V and 4 °C for 2 hours. Binding reactions were added appropriately to demonstrate a protein titration from zero to an appropriate concentration and run at 200 V and 4 °C for 45 min.

EMSA binding reaction setup

| Component | Volume (μL) |
|----------------------------------|-----------------------|
| 5 x EMSA Binding buffer | 2 |
| Fluorescently labelled probe DNA | Variable (250 fmoles) |
| BSA | 0.1 |
| Poly(d[IC]) | 1 |
| Protein of interest | 1 |
| dH ₂ O | to 10 μL |

2.8.1 EMSA visualisation and data processing

Three EMSA gels per assay were visualised using the Amersham Biosciences Typhoon 9400 on variable mode image in fluorescence mode, emission filter 526 SP. Sensitivity was set to normal. Band intensities were calculated using the grid scan feature and triplicate data processed in Prism (GraphPad Software) to create graphs seen in the results section. Fractional saturation curves were produced with fractional saturation, Y, varying from 0 – 1.0. Y values are calculated by $(Y/(Y+(1-Y)))$ and plotted against protein concentration. Data were converted to the Hill plot to analyse the degree of cooperativity in the binding events, characterised by the Hill coefficient (slope of the plot at $\log(\theta)=0$). The Hill plot is constructed by plotting $\log\theta$ against $\log[\text{protein}]$, with θ defined as $(\theta = (Y/(1-Y)))$. Dissociation coefficients (K_d) can also be extracted from the Hill plot as $K_d = 10^{\text{X-intercept}}$.

2.9 Protein crystallisation

2.9.1 Crystallisation screens

The protein AbiEi, once at optimum purity, was dialysed into crystallisation buffer and concentrated to 12 mg/mL. Commercially available 96-well crystallisation screens (Molecular Dimensions) were used to assay preferred crystallisation conditions for AbiEi. Sitting drop crystallisation trials were set-up using an Innovadyne Screenmaker robot. Table 2.8 shows the screens tested for AbiEi and their respective temperature for crystallisation. Each condition was tested at 8 mg/mL and 6 mg/mL by altering the ratio of protein to mother-liquor on the robotic crystallisation stage.

2.9.2 Crystallisation screen optimization

A customized crystallisation screen was designed to further test a range of conditions based on two early identified positive conditions. A range of conditions were designed to optimise conditions present in JCSG-*plus* (condition H3), and Pact Premier (condition F10). Both presented the same buffer, salt and precipitant. The customised screen tested a range of buffer pH, salt concentration, and PEG concentration.

2.9.3 Crystal harvesting

Positive individual crystallisation conditions were identified by microscopy. Mother liquor from the appropriate condition and 100% glycerol were mixed in a ratio of 1:1 and an equal volume of this mixture was added to the sitting drop. The crystal was visualised under a microscope to check for durability or deformity at this point. Surviving crystals were harvested using cryo-loops of appropriate sizes and snap-frozen in liquid nitrogen before adding to a puck submerged in liquid nitrogen for storage and transport.

2.10 X-ray crystallography

2.10.1 Data collection

Data collection was performed at Diamond Light Source, Oxford, UK, via remote access on beamline I03. Data sets were collected from a number of native AbiEi protein crystals. These datasets were merged in downstream processing.

2.10.2 Data processing

Initial data processing was automated by Diamond Light Source iSpyB using the X-ray image integration programs Xia2 and Xia2-DIALS (Winter and IUCr, 2010). Image integration and space group selection were carried out manually using the same programs as well as Mosflm (Leslie and Powell, 2007) to confirm maximum resolution and space group. Data reduction and generation of R sets was carried out using AIMLESS (Evans et al., 2013). Molecular replacement was run using PHASER (McCoy and IUCr, 2007) using Rv2827c (PDB code: 1ZELA (Janowski et al., 2009)) as a model. Following PHASER, refinement was carried out in REFMAC5 (Murshudov et al., 2011) or via rigid body refinement in PHASER. ARCIMBOLDO (Rodríguez et al., 2012) was used for *ab initio* phasing and chain tracing to position ideal molecular fragments using PHASER and SHELXE (Sheldrick and IUCr, 2008).

Table 2.8 Crystallisation screens for AbiEi

| Crystallisation screen | Product code ^a | Temperature |
|------------------------|---------------------------|-------------|
| BCS | MD1-105 | 18 °C |
| Clear Strategy 1 | MD1-31 | 18 °C |
| Cleat Strategy 2 | MD1-32 | 18 °C |
| JCSG- <i>plus</i> | MD1-40 | 18 °C |
| JCSG- <i>plus</i> | MD1-40 | 4 °C |
| LMB | MD1-99 | 18 °C |
| Morpheus | MD1-47 | 18 °C |
| NR-LBD + Extension | MD1-34 | 18 °C |
| Pact Premier | MD1-36 | 18 °C |
| Pact Premier | MD1-36 | 4 °C |
| SG1 | MD1-89 | 18 °C |
| Structure Screen 1 + 2 | MD1-30 | 18 °C |

^a – product codes can be used to find these commercially available screens at <https://www.moleculardimensions.com>

Chapter 3. Results

3.1 Antitoxin protein expression and purification

The *S. agalactiae* antitoxin protein, AbiEi, and *M. tuberculosis* antitoxin proteins, Rv1044 and Rv2827c, were expressed using the pET vector pSAT1-LIC via the T7 system in ER2566 following already published methods with minor amendments as appropriate (Blower et al., 2016). A combination of the purification techniques set out in 2.7 resulted in the purified proteins as visualised by SDS-PAGE in Figure 3.1. The shift in molecular weight from the tagged recombinant protein in the soluble fraction of the first nickel pull-down, to the isolated final protein sample at the appropriate molecular weight demonstrates successful cleavage of the affinity tag (Figure 3.1). It is worth noting that no detectable protein was made during trial expressions of Rv0837c. The *rv0837c* gene contains multiple rare codons and so an optimised expression sequence was synthesised and trialled for expression, but again no protein expression was detected. It was decided to continue the studies using purified AbiEi, Rv1044 and Rv2827c.

3.2 DNA-binding via EMSA

To test whether our type IV antitoxins are capable of binding DNA, EMSAs were performed using DNA sequences from the promoter regions of each system (detailed in Table 2.6). Our initial aim was to corroborate previously published data on the DNA-binding activity of AbiEi (Hampton et al., 2018).

3.2.1 AbiEi binds to a single inverted repeat in the *AbiEi* promoter and displays positive cooperativity

AbiEi has been shown to autoregulate expression of the *abiE* operon through positive cooperative binding of the antitoxin to inverted repeat sequences within and slightly upstream of the *abiE* promoter (Hampton et al., 2018). To validate our experimental system, DNA-binding studies by EMSA were repeated to confirm positive cooperative binding by AbiEi. AbiEi titrations against promoter region probes (as per Table 2.6, TRB1061 – TRB1064) demonstrated clear band shifts similar to already published data (Hampton et al., 2018) (Figure 3.2a-d). Densitometry of the EMSAs against a WT probe (Figure 3.2a) was performed and the data were used to make semi-

quantitative analyses of DNA-binding in the form of a saturation curve and Hill plot (Figure 3.2e-f). As the Hill co-efficient (or Hill slope) is significantly greater than 1 at 1.553 ± 0.31 AbiEi bound the WT probe in a positively cooperative manner (Figure 3.2f). This result matches published data (Hampton et al., 2018). Further saturation curves and Hill plots (Figure 3.3) were then generated for EMSAs containing a single knock-out of the inverted sequences (Figure 3.2b-c). As only the single binding site is present, no level of cooperativity could be established, however it is clear to see that affinity for the two binding sites, IR1 and IR2, is roughly equal with K_d 's of $0.593 \pm 0.003 \mu\text{M}$ and $0.607 \pm 0.005 \mu\text{M}$ respectively (Figure 3.3c and Figure 3.3e respectively).

Figure 3.1 Protein expression and purification

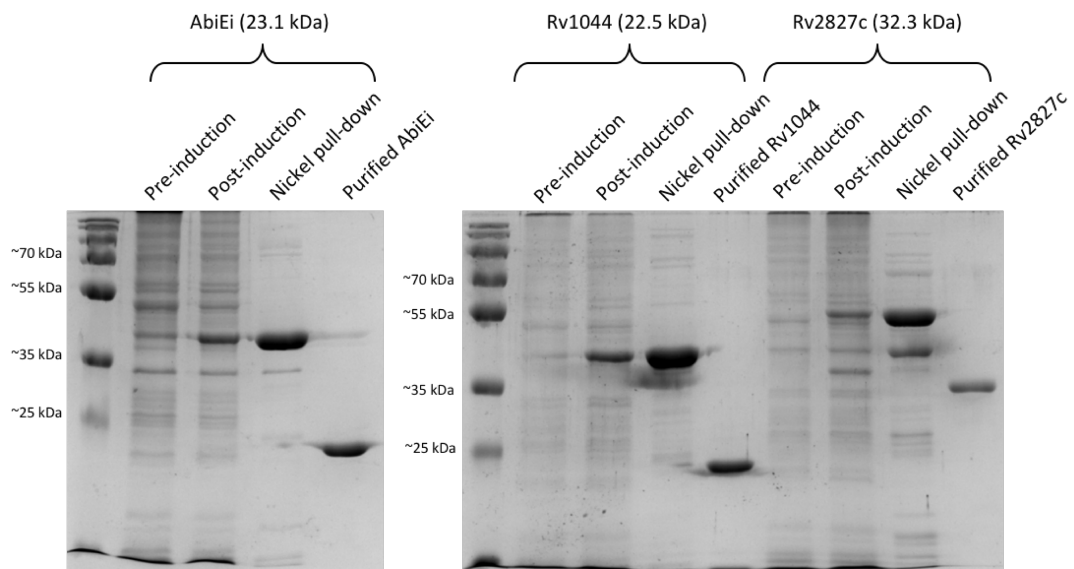


Figure 3.1. Pre- and Post- induction samples were taken before and after the addition of IPTG for protein expression. These samples were resuspended in appropriate amounts of buffer A500 and boiled at 95 °C to lyse the bacterial cells. Post-induction samples show selectively induced protein expression of our target proteins which can be successfully pulled down via the N-terminal 6-His tag and purified after cleavage of the 6-His-SUMO2 fusion (successful cleavage indicated by an approximately -11 kDa shift in molecular weight).

Figure 3.2 DNA-binding studies of AbiEi with *abiEi* -1 to -71

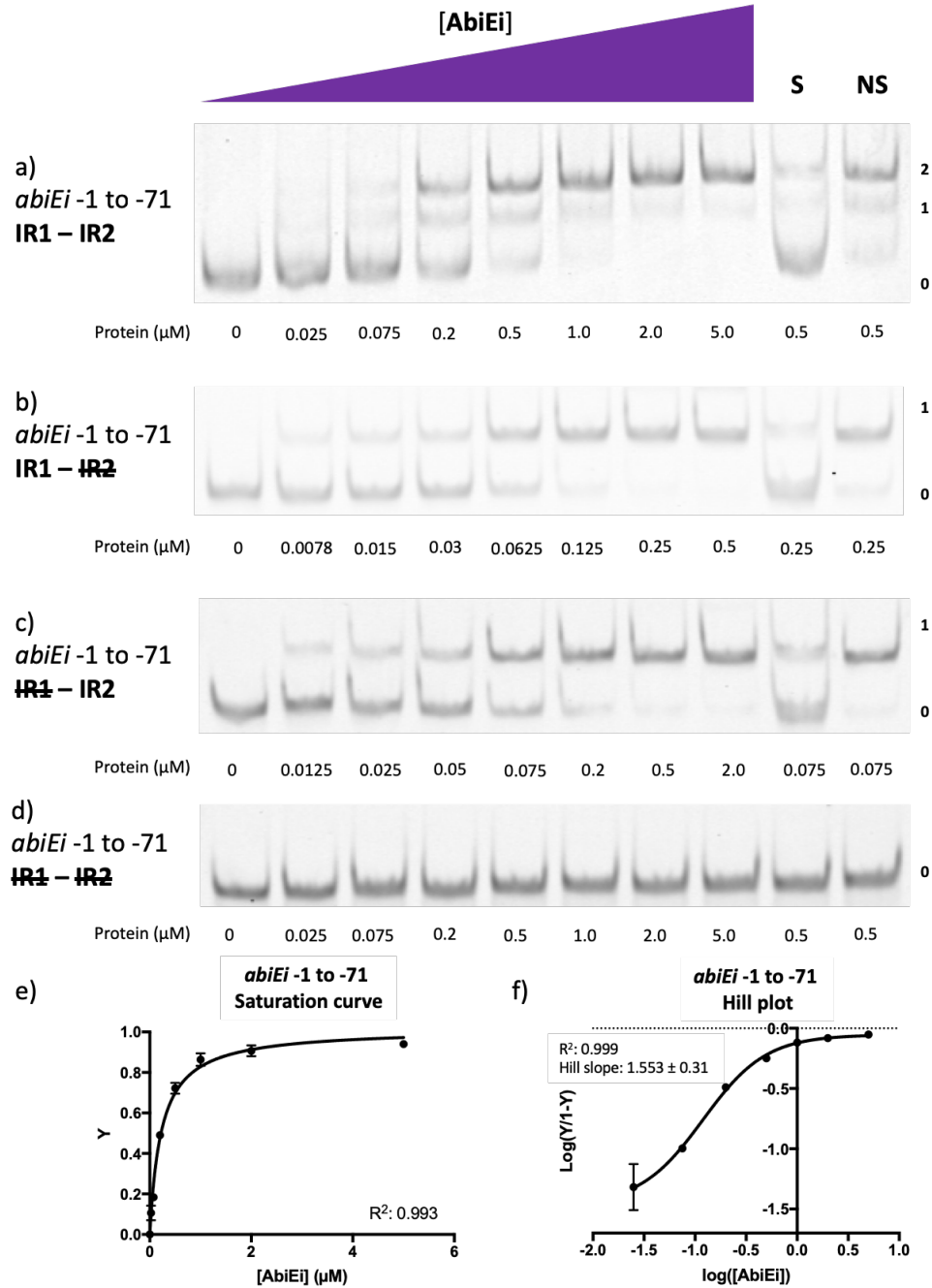


Figure 3.2. AbiEi titrations against 250 femto-moles fluorescently labelled dsDNA probe oligos. a) Oligo TRB1061 contains two inverted repeats (IR1 and IR2) intact; b) Oligo TRB1062 contains IR1 intact and IR2 substituted for a poly-C track; c) Oligo TRB1063 contains IR2 intact and IR1 substituted for a poly-C track; d) Oligo TRB1064 contains both inverted repeats, IR1 and IR2, substituted for a poly-C track; e) Fractional saturation curve plotted using the EMSA data seen in a); f) Hill plot using the data from the EMSA seen in a). Points are plotted from triplicate data and display mean values with SEM. S – each experiment contained 100-fold excess of the specific unlabelled probe; NS – each experiment contained 100-fold excess of non-specific unlabelled probe. Numbering -1 to -71 denotes the promoter region included in the probe, upstream of the transcriptional start site.

Figure 3.3 Saturation curves and Hill plots for individual inverted repeats of the *abiEi* promoter -1 to -71 region

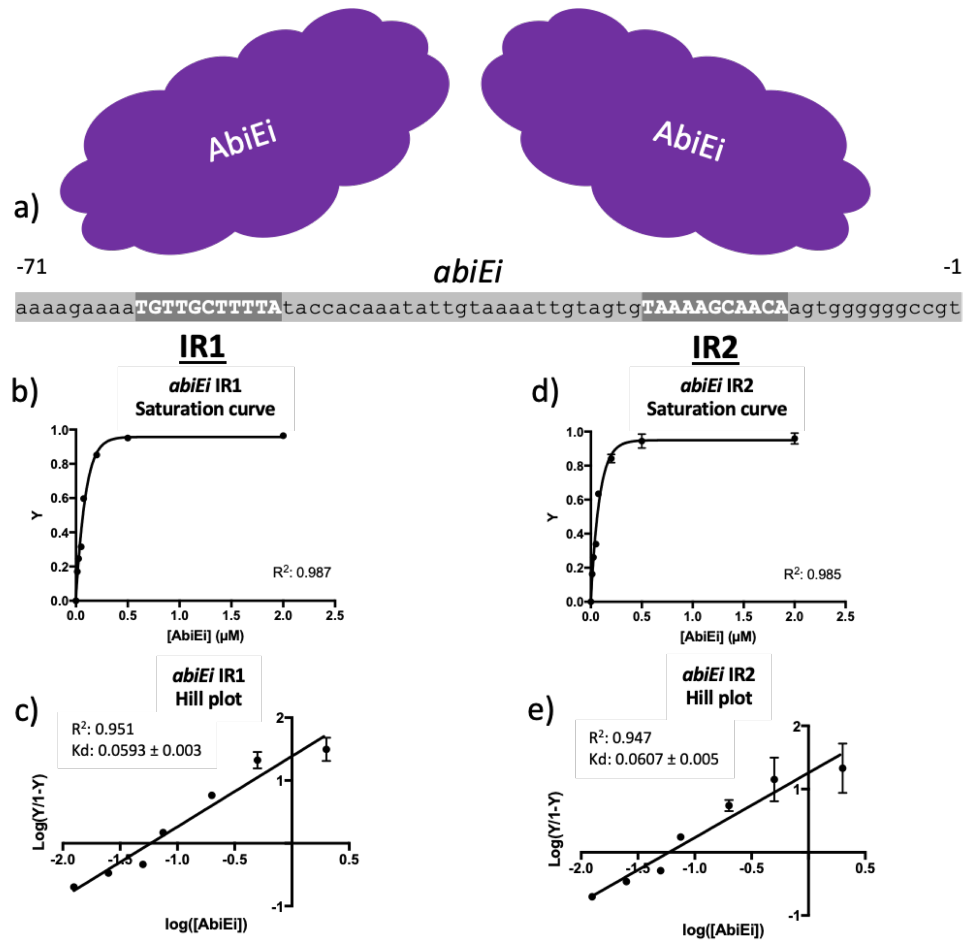


Figure 3.3. Saturation curves and Hill plots for individual repeats of the *abiEi* promoter -1 to -71 region. a) Schematic demonstrating the binding of AbiEi (structure unknown) to the *abiEi* -1 to -71 promoter region probe inverted repeats IR1 and IR2. b – c) Plots from Figure 3.2 b); d – e) plots from Figure 3.2 c). Points are plotted from triplicate data and display mean values with SEM. Hill plots display calculated K_d for each site ($K_d = 10^{\text{X intercept}}$). Numbering -1 to -71 denotes the promoter region included in the probe, upstream of the transcriptional start site.

3.2.2 Rv1044 binding site is yet to be identified

On testing Rv1044 against the 130 base pairs directly up-stream within the *rv1044* promoter region, no DNA-binding was seen (Figure 3.4). No shifts are evident despite a titration up to 5 μ M Rv1044 protein (Figure 3.4). Analysing the sequence of the regions tested in these assays highlighted no obvious inverted repeat or other DNA-binding sequence motifs. Bioinformatic analyses failed to highlight any candidate sequences for binding up to 500 bp upstream of the *rv1044* promoter.

3.2.3 Rv2827c binds to two inverted repeats in the *rv2827c* promoter

Bioinformatics was used to identify four 23 bp inverted repeats in the promoter region of the *rv2827c/rv2826c* operon between -1 and -131 bp (Figure 3.5 created using weblogo.berkeley.edu (Crooks et al., 2004)). These were labelled inverted repeats (IR) 1 to 4, with 1 being the furthest upstream of the transcriptional start and 4 being the closest. The repeats were included in two separate dsDNA oligos listed in Table 2.6 and assayed against Rv2827c, covering regions -1 to -71 and -61 to -131. These regions were first tested as two consecutive pairs to allow for direct comparison with the *AbiEi* promoter (Figure 3.3a).

3.2.3.1 Rv2827c binding to *rv2827c* -1 to -71 requires further study

Rv2827c bound to both inverted repeats within the -1 to -71 region, however saturation remained low as can be seen in the EMSA (Figure 3.6a) and corresponding saturation curve (Figure 3.6e). Sequential removal of the inverted repeats by mutating one, the other or both to poly-C tracts reduced Rv2827c to a single binding event (Figure 3.6b – c) or ablated binding (Figure 3.6d). This contrasts data for the binding of *AbiEi* to its cognate repeats where saturation occurs at lower protein concentrations (Figure 3.6e, cf Figure 3.2e). Despite the obviously low saturation, the Hill plot indicated positively cooperative binding with a Hill slope of 1.172 ± 0.52 . This result cannot be trusted due to the extremely large error associated with the Hill slope, and is therefore likely not a true reflection of the degree of cooperativity, if any, for binding to this region. This experiment requires repeating to obtain more accurate data and a better understanding of the protein-DNA interactions of this IR pair.

3.2.3.2 Rv2827c binds in a negatively cooperative manner to *rv2827c* -61 to -131

Similarly, Rv2827c bound to both inverted repeats within the -61 to -131 region (Figure 3.7a) however for this region, saturation appears to have increased (Figure 3.7e) in comparison to that seen for the -1 to -71 region (Figure 3.7e, cf Figure 3.6e) but remains low in comparison to AbiEi with its cognate promoter (Figure 3.7e, cf Figure 3.2e). Again, sequential removal of the inverted repeats by mutating one, the other or both to poly-C tracts reduced Rv2827c to a single binding event (Figure 3.7b – c) or ablated binding (Figure 3.7d). Data analysis via the Hill plot indicates that the binding to this region is negatively cooperative with a Hill slope of 0.846 ± 0.07 (Figure 3.7f). This is a surprising result having initially predicted a similar degree of cooperativity seen for AbiEi (Figure 3.2e).

Figure 3.4 DNA-binding studies of Rv1044 with *rv1044* -1 to -131

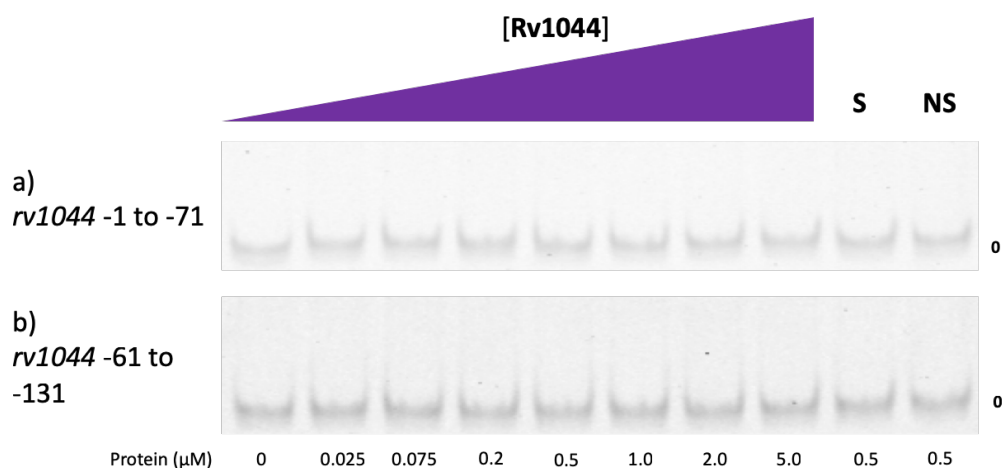


Figure 3.4 Rv1044 titrations against 250 femto-moles fluorescently labelled dsDNA probe oligos. a) Oligo TRB1086; b) Oligo TRB1102. Neither oligo sequence contains a recognised DNA-binding sequence motif. S – each experiment contained 100-fold excess of the specific unlabelled probe; NS – each experiment contained 100-fold excess of non-specific unlabelled probe. Numbering -1 to -131 denotes the promoter region included in the probe, upstream of the transcriptional start site.

Figure 3.5 Consensus motif for *rv2827c* inverted repeats 1 – 4

Inverted repeat sequence

Consensus motif

1. Acttgagtttgaacgggtttcg
2. Acttgatttaggaacgggcttcg
3. Tcttgagtttgaacattgttgcg
4. Acttgaattcacaccggaacta



Figure 3.6 DNA-binding studies of Rv2827c with *rv2827c* -1 to -71

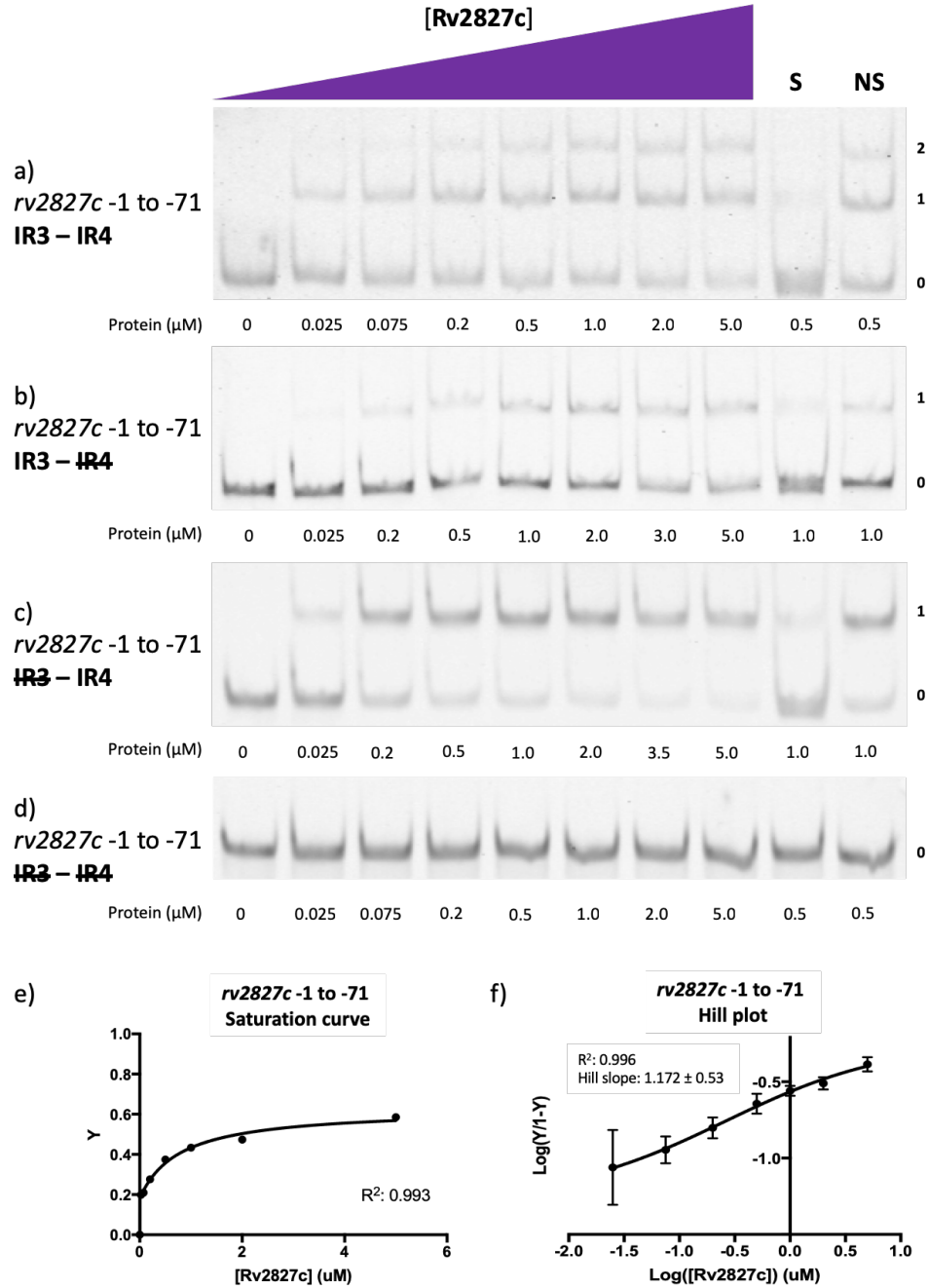


Figure 3.6. Rv2827c titrations against 250 femto-moles fluorescently labelled dsDNA probes for IR3 and IR4. a) Oligo TRB1104 contains two inverted repeats (IR3 and IR4) intact; b) Oligo TRB1273 contains IR3 intact and IR4 substituted for a poly-C track; c) Oligo TRB1271 contains IR4 intact and IR3 substituted for a poly-C track; d) Oligo TRB1274 contains both inverted repeats, IR3 and IR4, substituted for a poly-C track; e) Fractional saturation curve plotted using the EMSA data seen in a); f) Hill plot using the data from the EMSA seen in a). Points are plotted from triplicate data and display mean values with SEM. S – each experiment contained 100-fold excess of the specific unlabelled probe; NS – each experiment contained 100-fold excess of non-specific unlabelled probe. Numbering -1 to -71 denotes the promoter region included in the probe, upstream of the transcriptional start site.

Figure 3.7 DNA-binding studies of Rv2827c with *rv2827c* -60 to -131

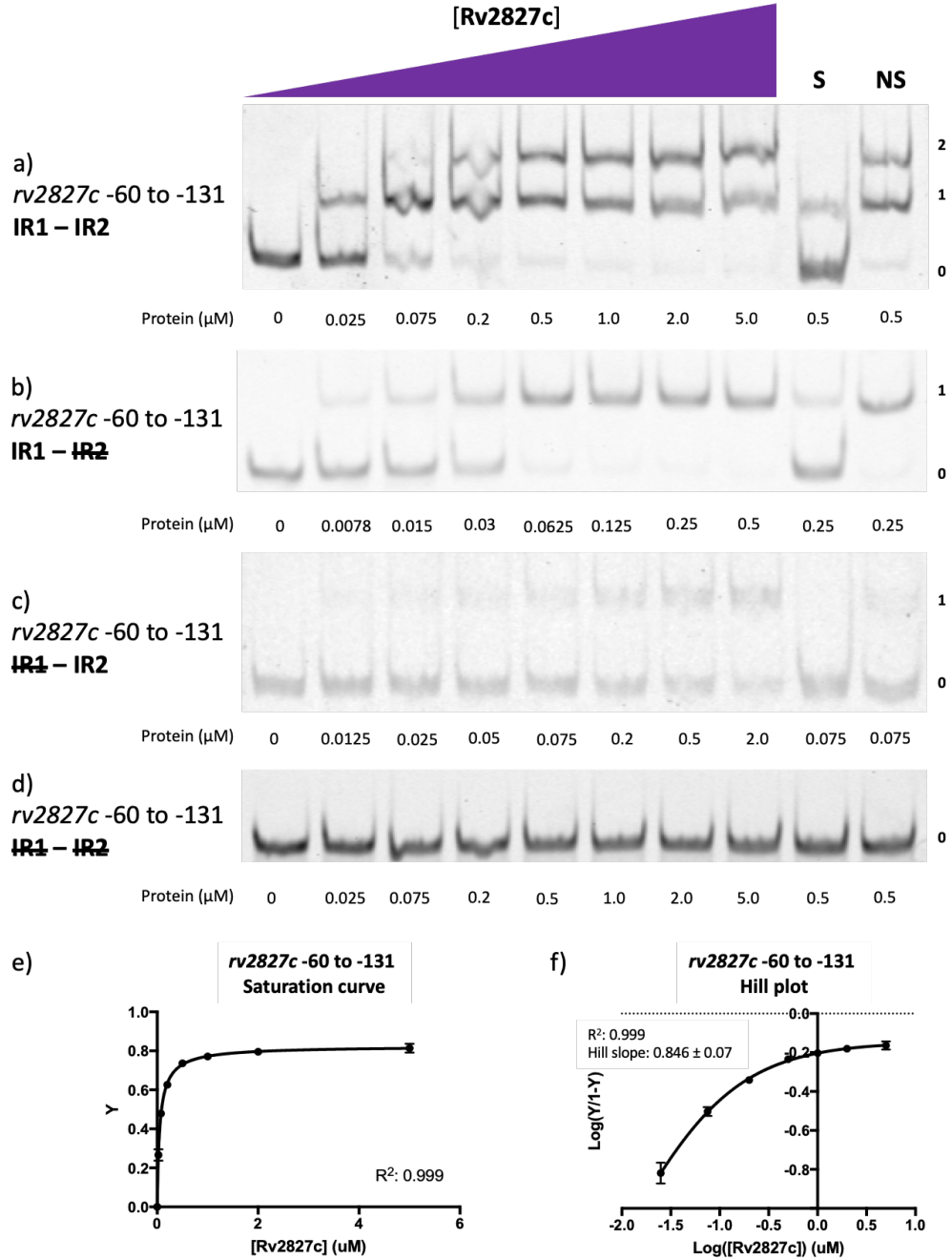


Figure 3.7. Rv2827c titrations against 250 femto-moles fluorescently labelled dsDNA probe oligos for IR1 and IR2. a) Oligo TRB1106 contains two inverted repeats (IR3 and IR4) intact; b) Oligo TRB1277 contains IR1 intact and IR2 substituted for a poly-C track; c) Oligo TRB1275 contains IR2 intact and IR1 substituted for a poly-C track; d) Oligo TRB1278 contains both inverted repeats, IR1 and IR2, substituted for a poly-C track; e) Fractional saturation curve plotted using the EMSA data seen in a); f) Hill plot using the data from the EMSA seen in a). Points are plotted from triplicate data and display mean values with SEM. S – each experiment contained 100-fold excess of the specific unlabelled probe; NS – each experiment contained 100-fold excess of non-specific unlabelled probe. Numbering -60 to -131 denotes the promoter region included in the probe, upstream of the transcriptional start site.

3.2.3.3 Rv2827c binds in a negatively cooperative manner to the -1 to -131 promoter region of *rv2827c*

A full-length probe covering the *rv2827c* region -1 to -131 was generated in order to examine the interaction of Rv2827c protein with all four identified inverted repeats. Four distinct species were observed, indicating that all four inverted repeats can be bound simultaneously by Rv2827c (Figure 3.8). Saturation of these sites was low (Figure 3.8b) and the Hill coefficient of 0.779 ± 0.12 confirmed negatively cooperative binding of Rv2827c across these four inverted repeats (Figure 3.8c). Displaying the saturation curve data on a semi-log scale (Figure 3.8d) highlights breaks and multiple distinct gradients in the binding curve, eluding to multiple individual binding events.

3.2.3.4 Negative cooperativity within the *rv2827c* -1 to -131 promoter region is DNA-sequence based rather than protein-protein interaction based

To analyse the protein-DNA interaction further and validate the negatively cooperative binding of Rv2827c, a perfect inverted repeat probe was created using two copies of IR4. IR4 was selected as the high affinity sequence in the IR pair with the greatest degree of negative cooperativity (Figure 3.6f). The perfect inverted repeat probe was made by maintaining the spacing between the repeats and IR3 was mutated to an exact inverted repeat of IR4. Using the perfect IR4 repeat probe, the Rv2827c-DNA interaction appeared stronger (Figure 3.9a). Data analysis showed higher occupancy of binding and a Hill coefficient of 0.945 ± 0.21 (Figures 3.9b-c), indicating Rv2827c interacts with each of the inverted repeats independently of each other in a more non-cooperative manner. Therefore, we can postulate that it is the DNA sequence of the IR that is causing the 'pseudo' negatively cooperative binding rather than protein-protein interactions.

To test this hypothesis, saturation curves and Hill plots were generated for each independent IR (Figure 3.10). The resulting K_d values support the observed rapidity of binding saturation for IR1 and IR4, and the slow binding saturation of IR2 and IR3. Ranking the IRs by K_d , we can suggest a model for promoter binding wherein Rv2827c will bind first to IR1, closely followed by IR4, then IR2 and finally IR3.

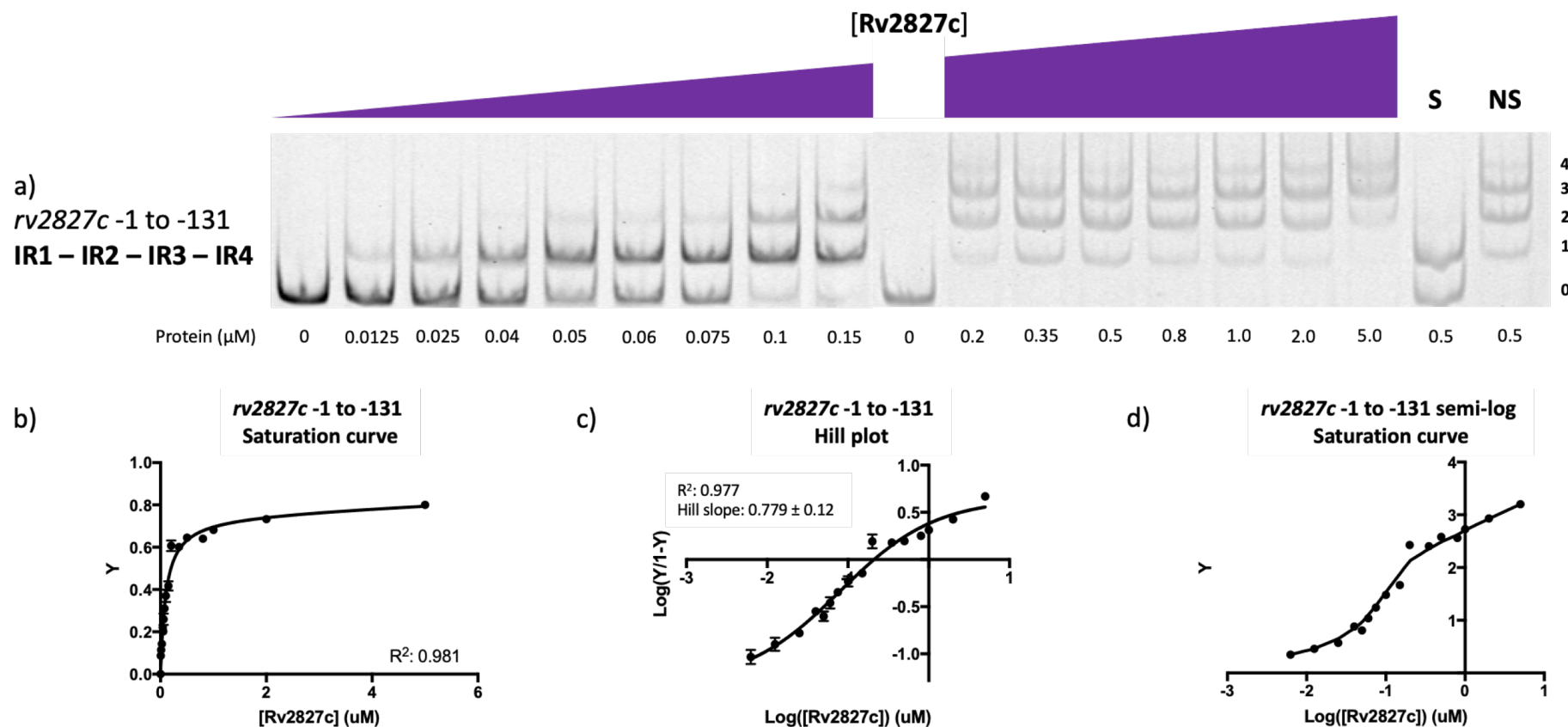
Figure 3.8 DNA-binding studies of Rv2827c with *rv2827c* -1 to -131

Figure 3.8. DNA-binding studies of Rv2827c with *rv2827c* -1 to -131. a) EMSA image from the Rv2827c titration against 250 femto-moles fluorescently labelled dsDNA probe oligo QS1 (*rv2827c* -1 to -131) with all four inverted repeats present; b) Fractional saturation curve plotted using the EMSA data seen in a); c) Hill plot using the EMSA data seen in a); d) Semi-log saturation curve plotted using the EMSA data seen in a) showing distinct breaks in the binding pattern, in accordance with the multiple binding sites present in the probe. Points are plotted from triplicate data and b) displays mean values with SEM. S – each experiment contained 100-fold excess of the specific unlabelled probe; NS – each experiment contained 100-fold excess of non-specific unlabelled probe. Numbering -1 to -131 denotes the promoter region included in the probe, upstream of the transcriptional start site.

Figure 3.9 DNA-binding studies of Rv2827c with *rv2827c* -1 to -71 IR4 perfect repeat

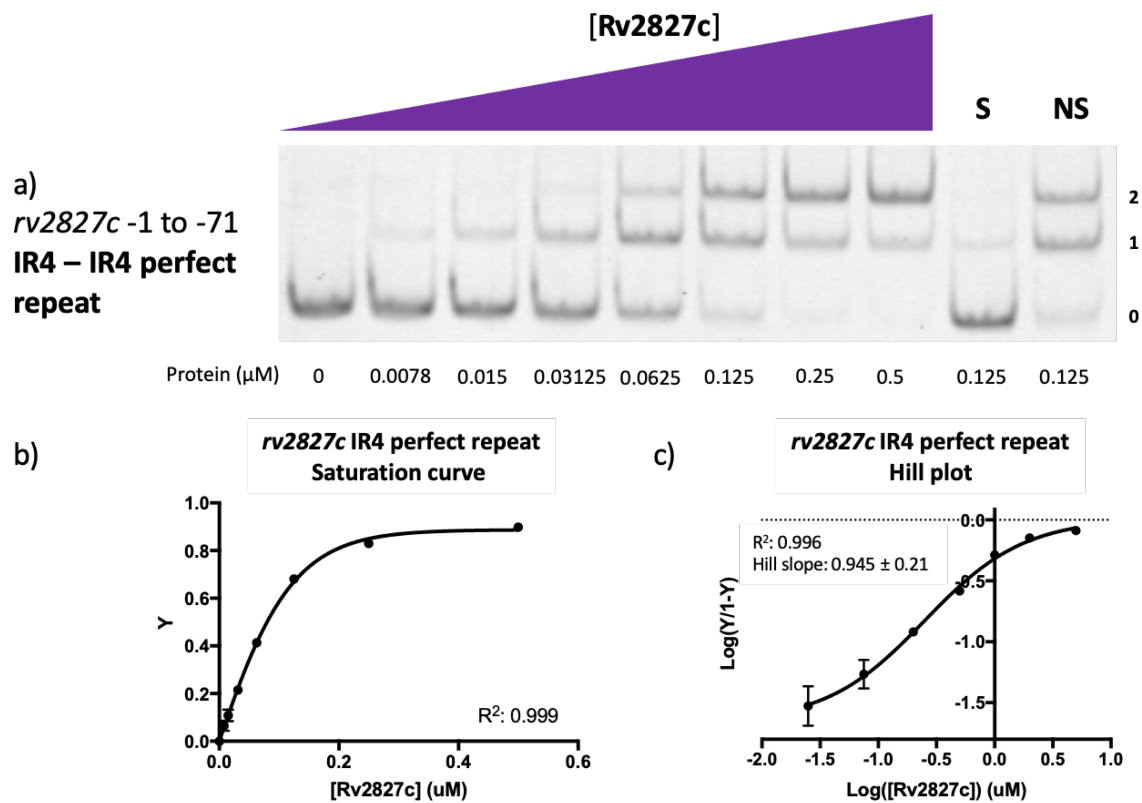


Figure 3.9. DNA-binding studies of Rv2827c with *rv2827c* -1 to -71 IR4 perfect repeat. a) EMSA with oligo TRB1296 containing IR3 substituted for a perfect inverted repeat of IR4; b) Fractional saturation curve plotted using the EMSA data seen in a); c) Hill plot using the data from the EMSA seen in a). Points are plotted from triplicate data and display mean values with SEM. S – each experiment contained 100-fold excess of the specific unlabelled probe; NS – each experiment contained 100-fold excess of non-specific unlabelled probe. Numbering -1 to -71 denotes the promoter region included in the probe, upstream of the transcriptional start site.

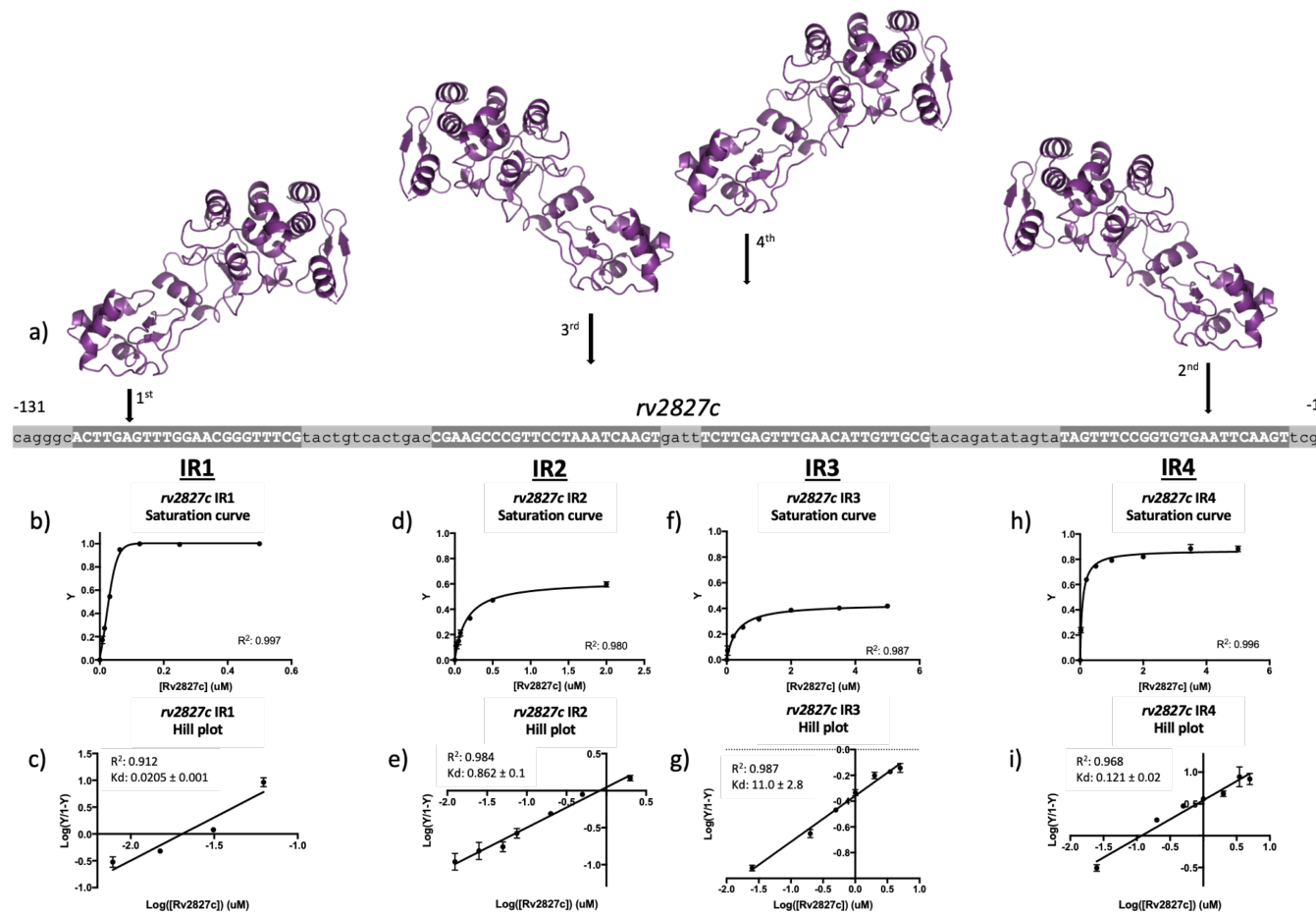
Figure 3.10 Saturation curves and Hill plots for individual inverted repeats of the *rv2827c* promoter -1 to -131 region

Figure 3.10 Legend overleaf

Figure 3.10. Saturation curves and Hill plots calculated from EMSAs examining binding of individual inverted repeats. a) Schematic demonstrating Rv2827c (PDB: 1ZELA – Janowski *et al.*, 2009) binding to the *rv2827c* upstream region in order as determined by calculated Kds for each IR. Plots from Figure 3.5; e – h) Plots from Figure 3.6. Points are plotted from triplicate data using mutant probes and display mean values with SEM. Hill plots display calculated Kd for each site ($K_d = 10^{X_{\text{intercept}}}$). Proteins are cartoon structures of Rv2827c (PDB: 1ZELA – Janowski *et al.*, 2009) and are labelled with predicted order of binding.

3.2.4 Binding of type IV antitoxin proteins to non-cognate type IV toxin-antitoxin promoter regions

3.2.4.1 Rv1044 can bind to the *abiEi* WT promoter region IRs

Initially, in order to test the binding capabilities of Rv1044, once no *rv1044* promoter sequence could be identified for cognate binding, the *abiEi* -1 to -71 probe was used as a positive DNA-sequence control. In this assay Rv1044 successfully bound to the probe DNA resulting in a double shift in the EMSA gel (Figure 3.11a), similarly to AbiEi (Figure 3.2). Interestingly, with an inverted repeat known to be capable of facilitating positive cooperativity (seen in the AbiEi titrations), the binding of Rv1044 appeared to be non-cooperative as indicated by a Hill coefficient of 0.964 ± 0.26 , however, given the error this result will need validating by further repeats (Figure 3.11c).

3.2.4.2 Non-cognate promoter binding is unique to Rv1044-*abiEi*

To test whether the other antitoxins being examined were also capable of binding to non-cognate promoter regions, EMSAs were run using each antitoxin against the identified probes to which binding has been seen (Figure 3.12). Binding is not apparent between the remaining combinations of protein and type IV TA system promoter regions (Figure 3.12). Therefore, from our current data, we can say that the binding of Rv1044 to the *abiEi* promoter inverted repeat is unique amongst these systems. This interaction is also almost certainly a novelty with no functional significance considering the interspecies nature of this binding; Rv1044 being *M. tuberculosis* derived, and *abiEi* being of *S. agalactiae* origin.

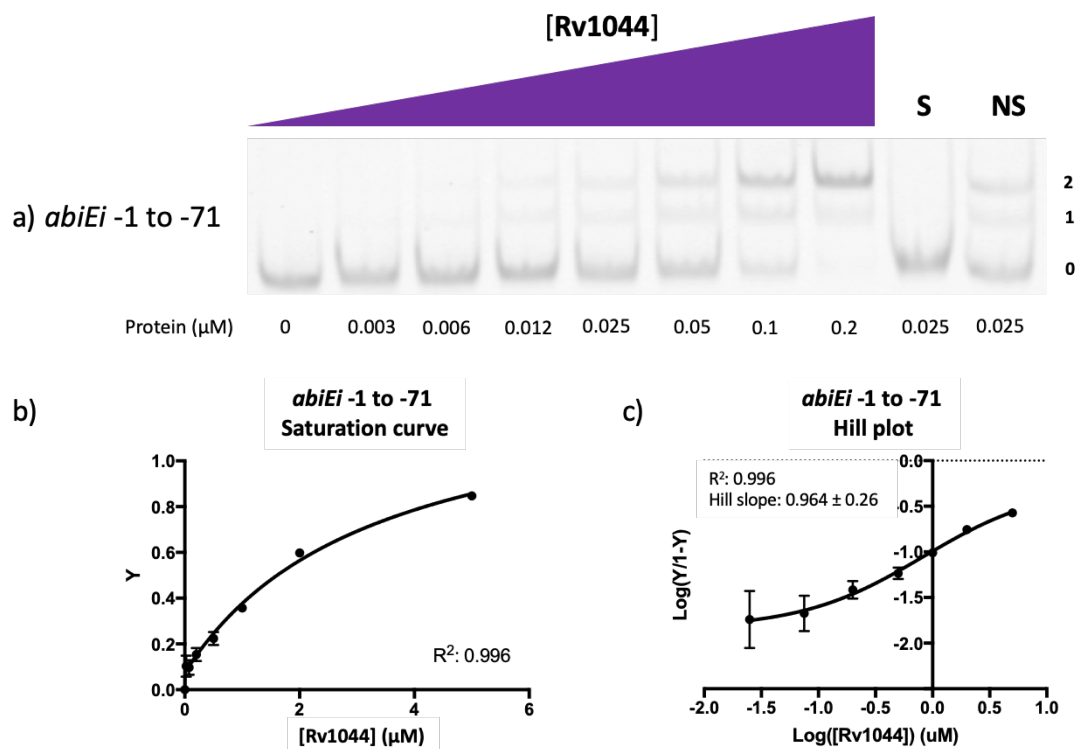
Figure 3.11 DNA-binding studies of Rv1044 with *abiEi* -1 to -71

Figure 3.11. DNA-binding studies of Rv1044 with *abiEi* -1 to -71. a) EMSA of Rv1044 with oligo TRB1061 containing the wild-type *AbiEi* promoter -1 to -71 region containing the intact inverted repeat sequences; b) Fractional saturation curve plotted using the EMSA data seen in a); c) Hill plot using the data from the EMSA seen in a). Points are plotted from triplicate data and display mean values with SEM. S – each experiment contained 100-fold excess of the specific unlabelled probe; NS – each experiment contained 100-fold excess of non-specific unlabelled probe. Numbering -1 to -71 denotes the promoter region included in the probe, upstream of the transcriptional start site.

Figure 3.12 DNA-binding studies of protein antitoxins with non-cognate promoter region probes

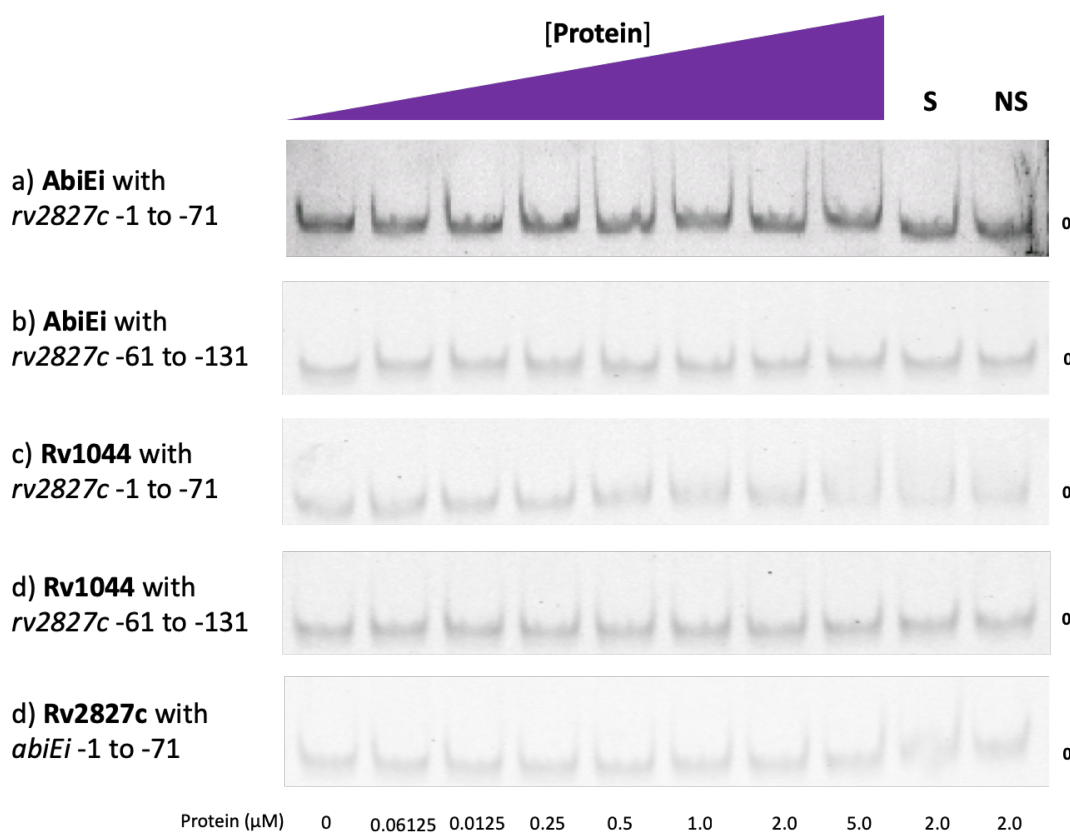


Figure 3.12. DNA-binding studies of protein antitoxins with non-cognate promoter region probes. a) EMSA with oligo TRB1296 containing IR3 substituted for a perfect inverted repeat of IR4; b) Fractional saturation curve plotted using the EMSA data seen in a); c) Hill plot using the data from the EMSA seen in a). Points are plotted from triplicate data and display mean values with SEM. S – each experiment contained 100-fold excess of the specific unlabelled probe; NS – each experiment contained 100-fold excess of non-specific unlabelled probe. Numbering -1 to -71 and -61 to -131 denotes the promoter region included in the probe, upstream of the transcriptional start site.

3.3 Crystallography

The protein AbiEi presented itself as an ideal crystallography candidate, having expressed at around 10 mg/L and subsequently been shown to be stable in solution at very high concentrations (> 25 mg/mL). Obtaining the crystal structure of AbiEi will help explain the functional data previously published on AbiEi (Hampton et al., 2018), provide further support for the activity of this antitoxin family, and allow direct comparisons between AbiEi and the structure of Rv2827c. Multiple screens were set as previously described, with varied protein concentrations and temperatures.

3.3.1 Crystallisation

AbiEi crystallised in multiple conditions with various buffers, salts, and precipitants. The crystal form was similar in each crystal ‘hit’ and returned on re-screening in the same conditions. The crystal shape is long and thin; many formed in needle shapes and in wider planes. Examples of the crystals can be seen in Figure 3.13 alongside their respective crystallisation condition.

3.3.2 X-ray crystallography data collection

Multiple data sets were collected at Diamond Light Source. An example of the diffraction pattern can be seen in Figure 3.13. All data sets refined to space group P1. Merging the best data sets resulted in a final resolution of 1.83 Å. Table 3.1 contains the data collection processing and statistics resulting from the merging of three independent datasets collected from three different native AbiEi crystals. The statistics from the data collection show high resolution data for protein crystals, with a high signal-to-noise ratio as indicated by the overall I/σ value.

3.3.2.1 Structure solving

Using the best data set obtained for AbiEi the structure could not be solved by molecular replacement using the structure of Rv2827c (PDB code: 1ZEL). Various attempts to modify and optimise the MR model using CHAINSAW and protein sequence alignment also resulted in failed MR. *Ab initio* structure solving could not be completed using the program ARCIMBOLDO. An experimental phasing method will need to be used to solve the AbiEi structure.

Figure 3.13 AbiEi crystals and diffraction pattern

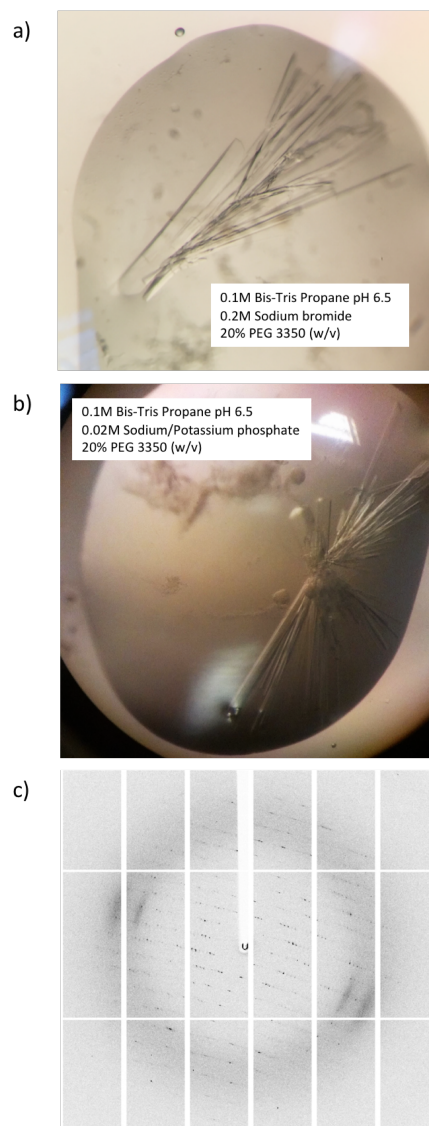


Figure 3.13. AbiEi crystals and diffraction pattern. a – b) Example crystals of AbiEi which regularly form long, thin planes; c) X-ray diffraction pattern from AbiEi protein crystals

Table 3.1 Data collection and refinement statistics

| Data Collection | AbiEi |
|---|-----------------------|
| Number of crystals | 3 |
| Beamline | iO3 |
| Space group | P1 |
| <i>Cell Dimensions</i> | |
| a, b, c (Å) | 34.24, 80.85, 122.17 |
| α, β, γ (°) | 102.48, 96.74, 100.47 |
| Wavelength (Å) | 0.97625 |
| Resolution low (Å) | 42.11 |
| Resolution high (Å) | 1.83 |
| Number unique reflections | 106618 (5213) |
| CC(1/2) | 0.999 (0.471) |
| Mean I/ σ I | 10.7 |
| Completeness (%) | 97.4 (96.1) |
| B-factor from Wilson plot (Å ²) | 28.99 |
| Values in parentheses are for the outer shell | |

Chapter 4. Discussion

4.1 Promoter region binding by type IV TA system antitoxins

The results presented so far appear to support the currently emerging theory that protein antitoxins of type IV TA systems are bifunctional and therefore similar to those of type II TA systems (Hampton et al., 2018). One function of each antitoxin is to relieve cellular stress as exerted by its toxin counterpart (without directly interacting); the other function is to regulate transcription of the TA system operon by promoter binding. The picture now emerges that, similar to type II systems, the overall activation and regulation of type IV TA systems is slightly more complex than originally thought and that autoregulation may also be a common and defining feature of type IV systems. Understanding the kinetics of the binding of type IV antitoxins with their respective promoters provides insights into the regulation of these systems. Calculating binding constants and analysing the nature of the interaction may allow us to better understand how exactly the systems can respond to counteract toxicity from their protein counterparts.

4.1.1 AbiEi-promoter interactions

The data for AbiEi binding to its cognate promoter region (Figure 3.3) demonstrated positively cooperative binding. This corroborated published work from our collaborators (Hampton et al., 2018) and demonstrated that our experimental system is reliable. However, values determined for the binding constants vary slightly between this study and published work; this may be due to differences in methods used to perform the EMSA, buffer differences, or sensitivity of equipment used to analyse band shifts. Nevertheless, the general trends remain the same and build confidence in our new data.

Positive cooperative binding is a key feature of the negative autoregulation observed in the AbiE system. Once one AbiEi protein has bound to an IR, this enhances the AbiEi protein binding to the second IR. Furthermore, assays have shown that AbiEi binding induces DNA bending (Hampton et al., 2018). The induced DNA-bending via protein binding to the recognised DNA sequences bring the IRs into closer spatial proximity to facilitate sequential binding (Khrapunov et al., 2006). AbiEi has been

shown to induce a DNA-bend of 72°, not only improving the spatial relationship of the inverted repeats, but also promoting AbiEi CTD interactions (Hampton et al., 2018). If the case is that two proteins bound to the promoter are required for full feedback inhibition, positive cooperative binding creates a rapid switch for shutting down transcription and therefore very tight regulation of the operon; hence, a rapidly responding system. In other words, once antitoxin levels reach a critical concentration determined by the binding constants, it is likely that transcription of both antitoxin and toxin will be repressed rapidly.

4.1.2 Mycobacterial antitoxin-promoter interactions

4.1.2.1 Rv1044 promoter interactions

The Rv1044 protein, despite being unable to bind the immediate region upstream of the transcriptional start site, does in fact possess the capability to bind DNA as seen by the binding to *abiEi* promoter sequences (Figure 3.11). This is likely due to their homology and a shared similar DNA-binding domain in structure, rather than sequence similarity. A possible reason for the negative cooperativity is that the amino acid-DNA base interactions that are favourable for AbiEi are not the same when Rv1044 binds. Structural similarity may have permitted the binding event; however, sequence dissimilarity dictated the affinity.

Such is the complexity of mycobacterial promoters, Rv1044 cognate binding motifs could be present further than 500 base pairs upstream from the transcriptional start site and still offer some functional significance in regulation of the operon when bound by protein. Unfortunately, none were identified even within 2000 bp upstream, using programs EMBOSS einverted (inverted repeat identifying software available at <http://emboss.bioinformatics.nl/cgi-bin/emboss/einverted>) and BLASTN (Altschul et al., 1990) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), searching for inverted repeats and aligned inverted repeats from already identified motifs, respectively. The question remains; does Rv1044 bind to its own promoter region and therefore support the emerging paradigm of type IV antitoxin autoregulation, or, is this protein anomalous and DNA-binding is simply a residual character to which the counterpart promoter DNA and necessity for autoregulation has been lost.

Some studies have highlighted a potential role for DNA-binding antitoxins acting as transcriptional regulators outside of their own systems in an emerging theory known as ‘cross-talk’ (Kasari et al., 2013; Walling and Butler, 2016; Wessner et al., 2015). Despite the theory of ‘cross-talk’ not being supported by our current analyses, this merely rules out a very small subset of promoters. The overall idea of protein antitoxins as part of a larger transcriptional network will be considered later in the context of all three protein antitoxins. To address this on a genome level would require genomics studies beyond the scope of this work.

4.1.2.2 Rv2827c promoter interactions

The structure of the Rv2827c protein has been solved to 1.93 Å and was predicted to have DNA-binding capabilities; the crystal structure includes a wHTH within the N-Terminal Domain (NTD) with three arginine-rich α -helices. This further supported the initial hypothesis for Rv2827c being a DNA-binding protein (Janowski et al., 2009).

The identification of candidate sequences using EMBLSS einverted within the *rv2827c* promoter initially returned the two sites denoted *rv2827c* IR1 and *rv2827c* IR2. Following a simple alignment of these two sequences with the 500 bp upstream region, the two sites denoted *rv2827c* IR3 and *rv2827c* IR4 were identified (these were initially missed using the default settings of einverted due to low completeness scores).

The data presented in this study (Figure 3.10) demonstrate a potentially novel mechanism for antitoxin-promoter binding in that four sites of differing Kds create a system mimicking negative cooperativity under these *in vitro* experimental conditions. A similar system has been reported for the binding of Cu²⁺ to a Prion protein whereby four distinct and varied Kds can be seen to contribute to negative cooperativity allowing the protein to respond to a wide range of Cu²⁺ concentrations; the physiological relevance has not, however, been agreed upon (Kozlowski et al., 2010).

Interestingly, our results show that the binding of Rv2827c to the *rv2827c* promoter region can be improved by manipulating the DNA sequence; Figure 3.9 shows a probe containing the IR4 exact inverted repeat return a Hill coefficient of 0.945 ± 0.21 . This value indicates non-cooperative binding and poses that the IR sequences are bound to independently of each other. To explore the possibility of protein-protein interaction facilitating positive-cooperativity perfect repeat pairs for each of IR1 - 4 should be tested, and most importantly, the strongest identified binding sequence, IR1. In theory, if Rv2827c can bind in a sequence dependent positive cooperative manner, IR1 would elicit the greatest positive cooperative response. Conversely, the weakest sequence, IR3, would in theory return very negatively cooperative results.

It will be useful to test the necessity and importance of the spacer between each inverted repeat and determine what part this short sequence plays in the efficiency of protein-DNA kinetics. It was shown for AbiEi that even when the space was increased to 50 bp some degree of positively cooperative binding was still apparent (Hampton et al., 2018). This may be a useful and informative assay to conduct on the native promoter sequence to check if the kinetics alter when the inverted repeat sequences remain in the same order. Furthermore, repeating spacer alterations for the positively cooperative binding model of the IR4 perfect repeat (as well as the IR1 perfect inverted repeat if positive cooperative results are obtained) should provide useful insight into DNA-bending in this system as well as positive cooperativity.

Negative cooperativity has been reported to be just as prevalent as positive cooperativity (Bush et al., 2012; Levitzki and Koshland, 1969). A system demonstrating negative cooperativity is, as one would expect, comparatively delayed in response; the system will however be responsive over a wider range of ligand (in this case Rv2827c protein) concentrations rather than the tight range seen in positively cooperative systems. The question is though, at what point in the saturation of the promoter region is negative autoregulation and repression of the operon achieved. If we assume that all four inverted repeat regions are to be bound before the system is repressed, we can safely assume that this is unlikely to ever happen due to the high concentrations needed. However, if the repression is

somehow dose-responsive, in that the expression of the toxin and antitoxin genes are separately dependent on the number of sites bound we can now see a more complex system emerging.

The physical interactions documented here should be followed up and complemented by functional assays. The promoter of a system, or simply the short 70 – 140 bp sequence containing identified binding sites, can be cloned upstream of a reporter gene such as *lacZ*, in order to determine promoter expression levels through beta-galactosidase assays (Smale, 2010). Indeed, Rv2827c has been shown to be negatively autoregulatory using this reporter system with 500 bp directly upstream of *rv2827c* used as the promoter (Usher, unpublished). Following this positive result up with a refined promoter region, that is, the sequence containing all four IRs as shown in Figure 3.10 should provide better insight as to the functionality of our identified DNA-binding region. It will be important to perform a series of knock-out experiments on this region to highlight which IR, or combination of IRs, is important for negative autoregulation.

We could hypothesize that occupancy of IR1 and possibly IR4 as the initial binding sites may create a loop only fully repressing the toxin gene; low levels of antitoxin are transcribed, translated, bind to the promoter and turned over before transcripts for the toxin can be created due to the position of the antitoxin before the toxin. This theory may align with the ability of a negatively cooperative system to respond over a greater range of ‘inhibitor’ concentrations. This may maintain a steady background pool of antitoxin to antagonistically nullify toxic effects or to promote this cellular process without toxin presence, or even act as a transcription factor outside of its own system.

A tuneable antitoxin expression system in tandem with a dual quantifiable reporter mimicking the antitoxin and toxin genes under the *rv2827c* promoter could be used to analyse the nature of the negative autoregulation. This could then give us more information as to the response, that is how much of either gene is expressed, of the system to a given amount of antitoxin. Similar systems have been used previously in

yeast (Belli et al., 1998; Kainth et al., 2009) and plasmids are commercially available as dual reporters, however, not specifically set up to mimic a TA system. Figure 4.1 shows the envisaged molecular set up of this promoter assay.

Alternatively, another benefit of a negatively autoregulated system could be if competitor molecules are present, such as an additional transcription factor. Assuming the kinetics for binding this promoter remains the same for other molecules, such as a protein binding to increase transcription of the *rv2827c/rv2826c* operon, then the response will occur over a greater range of concentrations. This may be useful to limit expression overall.

Considering the documented importance of the *rv2827c* gene in *M. tuberculosis* growth and survival (Griffin et al., 2011; Keren et al., 2011; Sassetti et al., 2003; Torrey et al., 2016) it may be sensible to assume that the toxin product of *rv2826c* is incredibly potent and therefore Rv2827c is absolutely required to nullify it, or Rv2827c has an important role elsewhere in the cell. This may then support the hypothesis that the negative cooperativity has been adopted in this system to maintain a basal level of antitoxin expression when repressed, and a steady and useful level of antitoxin and toxin when activated.

On examination of the individual inverted repeat regions that make up the consensus motif, alongside the saturation data, we can attempt to explain the order of binding. We can deduce the importance of adenine in the first position from the weak binding seen for IR3. With IR1 and IR4 being the best binding sites we can also deduce that a purine base in position 7 is important, but guanine is favoured as seen in IR1. This is supported by the fact arginine residues (present in the DNA-recognition helix) favour guanine bases (Luscombe et al., 2001). This is then followed by IR2 with the thymine present at position 7. In order for this to hold true, the adenine in the premier position must be considered the most important, followed by the position 7 variations. It should be easy to analyse these hypotheses with a series of mutational studies using IR1, changing the base at positions 1 and 7 and assaying Rv2827c binding.

4.2 Crystallography

The structure of AbiEi could not be solved by molecular replacement using the structure of Rv2827c, probably due to the low sequence similarity of 11%. We can therefore assume the structures of these proteins are more dissimilar than originally thought. The *ab initio* structure solving program ARCIMBOLDO was used as AbiEi is predicted to contain multiple helices, however, this program was unable to provide any insight into the AbiEi structure. The merged dataset presented earlier is of very high resolution and incorporating the heavy element selenium into the protein structure using selenomethionine may allow us to solve the phase problem as has been completed in many studies (Hendrickson et al., 1990; Hunter et al., 2016; Paterson et al., 2006; Rice et al., 2000; Walden, 2010) (see additional information). Expressing and crystallising a selenomethionine labelled AbiEi protein, followed by single or multi- wavelength anomalous diffraction experiments should allow us to determine the position of the three methionine residues in the AbiEi structure and build out from there. As this is a well-documented and historically successful technique in macromolecular structure solving so is a logical next step. Alternatively, derivatisation of native protein crystals could be used, as was completed for Rv2827c (Janowski et al., 2009).

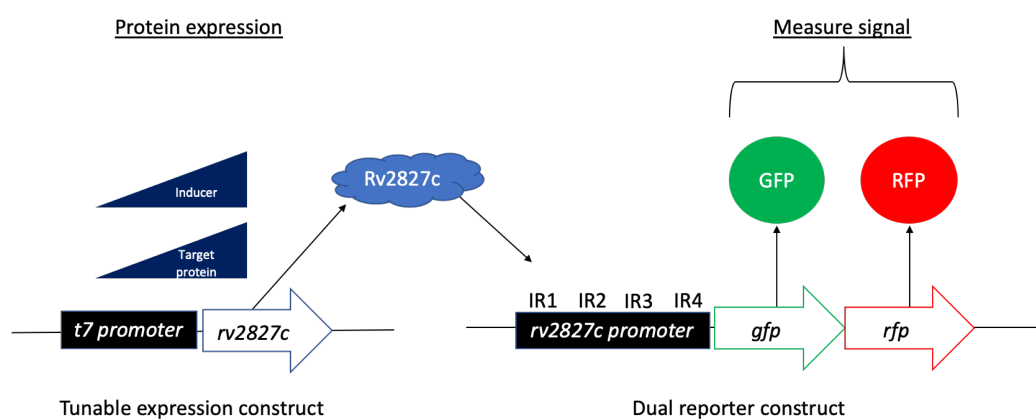
Figure 4.1 Schematic for *rv2827c* promoter assay with a dual reporter

Figure 4.1 Schematic for *rv2827c* promoter assay with dual reporter. A tuneable protein expression system (left) whereby the amount of Rv2827c produced can be controlled and estimated could be used alongside the dual reporter (right). Expressing a given amount of protein and measuring the signal of two reporter molecules, in this case two fluorescent proteins, with the knowledge of the order of binding from Figure 3.10 may provide insight into how the Rv2827c acts as an autoregulatory protein for the *rv2827c*/*rv2826c* operon.

4.3 Future directions

AbiEi has been studied relatively extensively when compared to the mycobacterial proteins analysed in this study. It has been shown that AbiEi fails to directly interact with its counterpart toxin AbiEii, yet is capable of neutralising the toxic effects. In order to properly characterise the mycobacterial antitoxins Rv1044 and Rv2827c, their affinity for cognate toxins Rv1045 and Rv2826c, respectively, should be analysed alongside their ability to counteract their partner toxin. Incubating purified samples of the toxin-antitoxin pairs together in an appropriate buffer followed by analytical FPLC using a calibrated superdex 200 increase column (GE Healthcare) could demonstrate an interaction by a chromatographic peak at the combined molecular weight, or lack thereof indicated by two separate peaks at their respective molecular weights. This technique can also be used to determine or confirm the monomeric or multimeric state of the proteins in solution. Toxicity assays and growth curves can be used to determine the effects on bacterial growth of the antitoxin alone, toxin alone, and co-expression to determine functionality as a TA pair. This will allow us to better understand the functional nature of the TA pairs and demonstrate whether growth can be rescued or not eluding as to whether the toxin is bacteriostatic or bactericidal respectively.

The toxin components should be tested for DNA-binding capabilities using the probes from this study containing binding motifs. DNA-binding studies using the same probes should also be conducted with the toxin and antitoxin present to assay whether interactions with DNA alter with the two components of the system present.

It is conceivable that antitoxins may have roles as transcriptional regulators outside of their own system. In order to better understand the antitoxins in this study and their full DNA-binding capabilities, DNA chip-seq can be performed (Kharchenko et al., 2008). This should identify the full range of DNA-sequences recognised and bound by our proteins. Following this with sequence alignment to the *M. tuberculosis* H37Rv genome may shed light on a transcriptional network for the protein antitoxins and identify possible binding sites for Rv1044 within the *rv1044* promoter. If promoters are identified, the beta-lactamase assay (Smale, 2010) can again be used to

determine whether the binding antitoxin is in-fact functional in regulating gene expression under that promoter. This data could then be analysed alongside *M. tuberculosis* transcriptomics, notably when the TA systems are active; it may be possible to piece together a functional network from here and demonstrate the complexity of the *M. tuberculosis* transcriptome during infection.

The expression and purification of Rv0837c is a necessary step to characterising all three putative type IV TA systems from *M. tuberculosis* and may be functionally the most important given the transcriptomic data presented earlier (Keren et al., 2011; Torrey et al., 2016). So far, this protein has not expressed in detectable amounts using *E. coli* with the native gene sequence, nor with a codon optimised sequence for *E. coli* expression. *Mycobacterium smegmatis* or *Mycobacterium thermoresistibile* may be more appropriate organisms being more closely related to express Rv0837c in. This could then enable the comprehensive analysis of the TA pairs as has been suggested previously.

Finally, as these systems are grouped together as part of the AbiEi abortive infective family of proteins, their effectiveness in aborting bacteriophage infection should be tested. These assays could also shed light on the native function of the systems: it is possible that they were initially incorporated to protect from bacteriophage infection but have now been adapted to respond to various stimuli including immune defence stress in the host. Cloning the entire locus for each system in to our model organism *E. coli* DH5 α and subjecting the transformed bacteria to *E. coli* phages may not provide useful information. Phages are highly specific to the bacteria they infect (Koskella et al., 2013) therefore, these systems logically should only respond to *M. tuberculosis* phage infection. It would be necessary to test phage-resistance in *Mycobacterium smegmatis*, or if available, *M. tuberculosis*.

In summary, the type IV antitoxins AbiEi, Rv1044, and Rv2827c can be expressed as soluble proteins and purified. These antitoxins have been shown capable of binding specific DNA sequences within promoter regions, demonstrating positive cooperativity in the case of AbiEi with its cognate promoter, and negative

cooperativity in the case of Rv2827c and its cognate promoter. Rv1044 can bind to the *S. agalactiae abiEi* promoter but not to any region yet identified in its cognate promoter. This highlights a structural relationship between Rv1044 and AbiEi.

AbiEi has been previously functionally characterised (Hampton et al., 2018) and was shown to be negatively autoregulatory, binding to two inverted repeats in the promoter region and our results corroborate this. The structural characterisation of AbiEi has begun and good-quality crystals have been generated, resulting in a current best data set of 1.83 Å resolution. Molecular replacement has not been possible with Rv2827c, indicating these proteins are less structurally related than initially believed (see additional information).

This is the first study in which the structurally characterised Rv2827c antitoxin protein has been shown to bind DNA, specifically to four inverted repeats in its cognate promoter. Interestingly, Rv2827c did so with negative cooperativity and we have shown that this is DNA-sequence dependent rather than protein-protein interaction dependent as generally accepted. We have therefore coined this term ‘pseudo-negative cooperativity’ as variations in the target binding sequence dictated the degree of cooperativity and these interactions ultimately occur independently of each other (as shown by mutational EMSAs creating a non-cooperative system). The promoter-binding events demonstrated here presents us with a potentially novel mechanism of negative autoregulation which requires further study. Understanding how each inverted repeat within the promoter relates to the function of negative autoregulation will improve our understanding of this essential gene and the TA system overall.

Additional information, post the research period

Having produced selenomethionine-derivatised AbiEi, the structure was solved in Jan 2019 by the MAD method. The global structure was similar to Rv2827c but the fold was sufficiently different to prevent reasonable alignments. This structure will be used for future studies.

Bibliography

Afif, H., Allali, N., Couturier, M., and Van Melder, L. (2001). The ratio between CcdA and CcdB modulates the transcriptional repression of the ccd poison-antidote system. *Mol. Microbiol.* **41**, 73–82.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Aslanidis, C., and de Jong, P.J. (1990). Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* **18**, 6069–6074.

Belli, G., Garí, E., Piedrafita, L., Aldea, M., and Herrero, E. (1998). An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast [published erratum appears in *Nucleic Acids Res.* 1998 Apr 1;26(7):following 1855]. *Nucleic Acids Res.* **26**, 942–947.

Berger, S.L., and Kimmel, A.R. (1987). *Guide to molecular cloning techniques*.

Blower, T.R., Fineran, P.C., Johnson, M.J., Toth, I.K., Humphreys, D.P., and Salmond, G.P.C. (2009). Mutagenesis and functional characterization of the RNA and protein components of the toxIN abortive infection and toxin-antitoxin locus of *Erwinia*. *J. Bacteriol.* **191**, 6029–6039.

Blower, T.R., Salmond, G.P., and Luisi, B.F. (2011a). Balancing at survival's edge: the structure and adaptive benefits of prokaryotic toxin–antitoxin partners. *Curr. Opin. Struct. Biol.* **21**, 109–118.

Blower, T.R., Pei, X.Y., Short, F.L., Fineran, P.C., Humphreys, D.P., Luisi, B.F., and Salmond, G.P.C. (2011b). A processed noncoding RNA regulates an altruistic bacterial antiviral system. *Nat. Struct. Mol. Biol.* **18**, 185–190.

Blower, T.R., Short, F.L., Rao, F., Mizuguchi, K., Pei, X.Y., Fineran, P.C., Luisi, B.F., and Salmond, G.P.C. (2012). Identification and classification of bacterial Type III toxin–antitoxin systems encoded in chromosomal and plasmid genomes. *Nucleic Acids Res.* **40**, 6158–6173.

Blower, T.R., Williamson, B.H., Kerns, R.J., and Berger, J.M. (2016). Crystal structure and stability of gyrase-fluoroquinolone cleaved complexes from *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 1706–1713.

Bøggild, A., Sofos, N., Andersen, K.R., Feddersen, A., Easter, A.D., Passmore, L.A., and Brodersen, D.E. (2012). The Crystal Structure of the Intact *E. coli* RelBE Toxin–Antitoxin Complex Provides the Structural Basis for Conditional Cooperativity. *Structure* **20**, 1641–1648.

Brantl, S. (2012). To cite this article: Sabine Brantl (2012) Bacterial type I toxin-antitoxin systems. *RNA Biol.* 9, 1488–1490.

Brennan, R.G., and Matthews, B.W. (1903). The Journal Of Biological Chemistry The Helix-Turn-Helix DNA Binding Motif.

Brennan, R.G., and Matthews, B.W. (1989). Structural basis of DNA-protein recognition. *Trends Biochem. Sci.* 14, 286–290.

Bush, E.C., Clark, A.E., DeBoever, C.M., Haynes, L.E., Hussain, S., Ma, S., McDermott, M.M., Novak, A.M., and Wentworth, J.S. (2012). Modeling the Role of Negative Cooperativity in Metabolic Regulation and Homeostasis. *PLoS One* 7, e48920.

Buts, L., Lah, J., Dao-Thi, M.-H., Wyns, L., and Loris, R. (2005). Toxin–antitoxin modules as bacterial metabolic stress managers. *Trends Biochem. Sci.* 30, 672–679.

Cataudella, I., Trusina, A., Sneppen, K., Gerdes, K., and Mitarai, N. (2012). Conditional cooperativity in toxin–antitoxin regulation prevents random toxin activation and promotes fast translational recovery. *Nucleic Acids Res.* 40, 6424–6434.

Chan, W.T., Espinosa, M., and Yeo, C.C. (2016). Keeping the Wolves at Bay: Antitoxins of Prokaryotic Type II Toxin-Antitoxin Systems. *Front. Mol. Biosci.* 3, 9.

van Crevel, R., Ottenhoff, T.H.M., and van der Meer, J.W.M. (2002). Innate immunity to *Mycobacterium tuberculosis*. *Clin. Microbiol. Rev.* 15, 294–309.

Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: A Sequence Logo Generator.

Crouch, T.H., and Klee, C.B. (1980). Positive Cooperative Binding of Calcium to Bovine Brain Calmodulin; Equilibrium dialysis measurements of the binding of Ca²⁺ to calmodulin have confirmed the existence of four high affinity Ca²⁺-binding sites (K_d between 3 X 10⁻⁶ and 2 X 10⁻⁶).

Dahlquist, F.W. (1978). [13] The meaning of scatchard and hill plots. *Methods Enzymol.* 48, 270–299.

Durand, S., Jahn, N., Condon, C., and Brantl, S. (2012). Type I toxin-antitoxin systems in *Bacillus subtilis*. *RNA Biol.* 9, 1491–1497.

Dy, R.L., Przybilski, R., Semeijn, K., Salmond, G.P.C., and Fineran, P.C. (2014). A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. *Nucleic Acids Res.* 42, 4590–4605.

Emond, E., Dion, E., Walker, S.A., Vedomuthu, E.R., Kondo, J.K., and Moineau, S. (1998). AbiQ, an abortive infection mechanism from *Lactococcus lactis*. *Appl. Environ. Microbiol.* **64**, 4748–4756.

Evans, P.R., Murshudov, G.N., and IUCr (2013). How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214.

Fineran, P.C., Blower, T.R., Foulds, I.J., Humphreys, D.P., Lilley, K.S., and Salmond, G.P.C. (2009). The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 894–899.

Franch, T., Gulyaev, A.P., and Gerdes, K. (1997). Programmed cell death by hok/sok of plasmid R1: Processing at the hok mRNA 3'-end triggers structural rearrangements that allow translation and antisense RNA binding. *J. Mol. Biol.* **273**, 38–51.

Garcia-Pino, A., Balasubramanian, S., Wyns, L., Gazit, E., De Greve, H., Magnuson, R.D., Charlier, D., van Nuland, N.A.J., and Loris, R. (2010). Allostery and Intrinsic Disorder Mediate Transcription Regulation by Conditional Cooperativity. *Cell* **142**, 101–111.

Goeders, N., Chai, R., Chen, B., Day, A., Salmond, G., Goeders, N., Chai, R., Chen, B., Day, A., and Salmond, G.P.C. (2016). Structure, Evolution, and Functions of Bacterial Type III Toxin-Antitoxin Systems. *Toxins (Basel)*. **8**, 282.

Goormaghtigh, F., Fraikin, N., Putrinš, M., Hallaert, T., Hauryliuk, V., Garcia-Pino, A., Sjödin, A., Kasvandik, S., Udekwu, K., Tenson, T., et al. (2018). Reassessing the Role of Type II Toxin-Antitoxin Systems in Formation of *Escherichia coli* Type II Persister Cells. *MBio* **9**, e00640-18.

Griffin, J.E., Gawronski, J.D., Dejesus, M.A., Ioerger, T.R., Akerley, B.J., and Sassetti, C.M. (2011). High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog.* **7**, e1002251.

Guglielmini, J., and Van Melderren, L. (2011). Bacterial toxin-antitoxin systems. *Mob. Genet. Elements* **1**, 283–306.

Guo, Y., Quiroga, C., Chen, Q., McAnulty, M.J., Benedik, M.J., Wood, T.K., and Wang, X. (2014). RalR (a DNase) and RalA (a small RNA) form a type I toxin-antitoxin system in *Escherichia coli*. *Nucleic Acids Res.* **42**, 6448–6462.

Gupta, A., Venkataraman, B., Vasudevan, M., and Gopinath Bankar, K. (2017). Co-expression network analysis of toxin-antitoxin loci in *Mycobacterium tuberculosis* reveals key modulators of cellular stress. *Sci. Rep.* **7**, 5868.

- Hampton, H.G., Jackson, S.A., Fagerlund, R.D., Vogel, A.I.M., Dy, R.L., Blower, T.R., and Fineran, P.C. (2018). AbiEi Binds Cooperatively to the Type IV abiE Toxin–Antitoxin Operator Via a Positively-Charged Surface and Causes DNA Bending and Negative Autoregulation. *J. Mol. Biol.* **430**, 1141–1156.
- Hendrickson, W.A., Horton, J.R., and LeMaster, D.M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**, 1665–1672.
- Hill, and Av. (1910). The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J Physiol* **40**, 4–7.
- Holden, D.W., and Errington, J. (2018). Type II Toxin-Antitoxin Systems and Persister Cells. *MBio* **9**, e01574-18.
- Hunter, M.S., Yoon, C.H., DeMirici, H., Sierra, R.G., Dao, E.H., Ahmadi, R., Aksit, F., Aquila, A.L., Ciftci, H., Guillet, S., et al. (2016). Selenium single-wavelength anomalous diffraction de novo phasing using an X-ray-free electron laser. *Nat. Commun.* **7**, 13388.
- Janowski, R., Panjekar, S., Eddine, A.N., Kaufmann, S.H.E., and Weiss, M.S. (2009). Structural analysis reveals DNA binding properties of Rv2827c, a hypothetical protein from *Mycobacterium tuberculosis*. *J. Struct. Funct. Genomics* **10**, 137–150.
- Kainth, P., Sassi, H.E., Peña-Castillo, L., Chua, G., Hughes, T.R., and Andrews, B. (2009). Comprehensive genetic analysis of transcription factor pathways using a dual reporter gene system in budding yeast. *Methods* **48**, 258–264.
- Kasari, V., Mets, T., Tenson, T., and Kaldalu, N. (2013). Transcriptional cross-activation between toxin-antitoxin systems of *Escherichia coli*. *BMC Microbiol.* **13**, 45.
- Kawano, M., Aravind, L., and Storz, G. (2007). An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. *Mol. Microbiol.* **64**, 738–754.
- Keren, I., Minami, S., Rubin, E., and Lewis, K. (2011). Characterization and transcriptome analysis of *Mycobacterium tuberculosis* persisters. *MBio* **2**, e00100-11.
- Kharchenko, P. V, Tolstorukov, M.Y., and Park, P.J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat. Biotechnol.* **26**, 1351–1359.
- Khrapunov, S., Brenowitz, M., Rice, P.A., and Catalano, C.E. (2006). Binding then bending: a mechanism for wrapping DNA. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19217–19218.

- Koskella, B., Meaden, S., Koskella, B., and Meaden, S. (2013). Understanding Bacteriophage Specificity in Natural Microbial Communities. *Viruses* 5, 806–823.
- Kozłowski, H., Łuczowski, M., and Remelli, M. (2010). Prion proteins and copper ions. Biological and chemical controversies. *Dalt. Trans.* 39, 6371.
- Leplae, R., Geeraerts, D., Hallez, R., Guglielmini, J., Drèze, P., and Van Melderen, L. (2011). Diversity of bacterial type II toxin–antitoxin systems: a comprehensive search and functional analysis of novel families. *Nucleic Acids Res.* 39, 5513–5525.
- Leslie, A.G.W., and Powell, H.R. (2007). Processing diffraction data with mosflm. (Springer, Dordrecht), pp. 41–51.
- Levitzi, A., and Koshland, D.E. (1969). Negative cooperativity in regulatory enzymes. *Proc. Natl. Acad. Sci. U. S. A.* 62, 1121–1128.
- Lewis, K. (2008). Multidrug Tolerance of Biofilms and Persister Cells. (Springer, Berlin, Heidelberg), pp. 107–131.
- Luscombe, N.M., Laskowski, R.A., and Thornton, J.M. (2001). Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level. *Nucleic Acids Res.* 29, 2860–2874.
- Madl, T., Van Melderen, L., Mine, N., Respondek, M., Oberer, M., Keller, W., Khatai, L., and Zangger, K. (2006). Structural Basis for Nucleic Acid and Toxin Recognition of the Bacterial Antitoxin CcdA. *J. Mol. Biol.* 364, 170–185.
- Maisonneuve, E., Castro-Camargo, M., and Gerdes, K. (2018). Erratum: (p)ppGpp Controls Bacterial Persistence by Stochastic Induction of Toxin–Antitoxin Activity (*Cell* (2013) 154(5) (1140–1150)(S0092867413009586)(10.1016/j.cell.2013.07.048)). *Cell* 172.
- Masuda, H., Tan, Q., Awano, N., Wu, K.-P., and Inouye, M. (2012). YeeU enhances the bundling of cytoskeletal polymers of MreB and FtsZ, antagonizing the CbtA (YeeV) toxicity in *Escherichia coli*. *Mol. Microbiol.* 84, 979–989.
- McCoy, A.J., and IUCr (2007). Solving structures of protein complexes by molecular replacement with *Phaser*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 63, 32–41.
- Milla, M.E., Brown, B.M., Waldburger, C.D., and Sauer, R.T. (1995). Shortle (1993) (Evans and Radford). Müller, P., Jahn, N., Ring, C., Maiwald, C., Neubert, R., Meißner, C., and Brantl, S. (2016). A multistress responsive type I toxin–antitoxin system: *bsrE* /SR5 from the *B. subtilis* chromosome. *RNA Biol.* 13, 511–523.

- Murre, C., McCaw, P.S., Vaessin, H., Caudy, M., Jan, L.Y., Jan, Y.N., Cabrera, C. V., Buskin, J.N., Hauschka, S.D., Lassar, A.B., et al. (1989). Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* 58, 537–544.
- Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., Vagin, A.A., and IUCr (2011). *REFMAC 5* for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67, 355–367.
- Oberer, M., Zangger, K., Gruber, K., and Keller, W. (2007). The solution structure of ParD, the antidote of the ParDE toxin-antitoxin module, provides the structural basis for DNA and toxin binding. *Protein Sci.* 16, 1676–1688.
- Overgaard, M., Borch, J., Jørgensen, M.G., and Gerdes, K. (2008). Messenger RNA interferase RelE controls *relBE* transcription by conditional cooperativity. *Mol. Microbiol.* 69, 841–857.
- Overgaard, M., Borch, J., and Gerdes, K. (2009). RelB and RelE of *Escherichia coli* Form a Tight Complex That Represses Transcription via the Ribbon–Helix–Helix Motif in RelB. *J. Mol. Biol.* 394, 183–196.
- Paterson, N.G., Riboldi-Tunncliffe, A., Mitchell, T.J., and Isaacs, N.W. (2006). Purification, crystallization and preliminary X-ray diffraction analysis of RafE, a sugar-binding lipoprotein from *Streptococcus pneumoniae*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 62, 676–679.
- Ramage, H.R., Connolly, L.E., and Cox, J.S. (2009). Comprehensive Functional Analysis of *Mycobacterium tuberculosis* Toxin-Antitoxin Systems: Implications for Pathogenesis, Stress Responses, and Evolution. *PLoS Genet.* 5, e1000767.
- Raumann, B.E., Brown, B.M., and Sauer, R.T. (1994). Major groove DNA recognition by β -sheets: the ribbon-helix-helix family of gene regulatory proteins. *Curr. Opin. Struct. Biol.* 4, 36–43.
- Rice, L.M., Earnest, T.N., Brunger, A.T., and IUCr (2000). Single-wavelength anomalous diffraction phasing revisited. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 56, 1413–1420.
- Rodríguez, D., Sammito, M., Meindl, K., de Ilarduya, I.M., Potratz, M., Sheldrick, G.M., and Usón, I. (2012). Practical structure solution with ARCIMBOLDO. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 68, 336–343.
- Sala, A., Bordes, P., and Genevaux, P. (2014a). Multiple toxin-antitoxin systems in *Mycobacterium tuberculosis*. *Toxins (Basel)*. 6.

- Sala, A., Bordes, P., and Genevau, P. (2014b). Multiple toxin-antitoxin systems in *Mycobacterium tuberculosis*. *Toxins (Basel)*. *6*, 1002–1020.
- Samson, J.E., Spinelli, S., Cambillau, C., and Moineau, S. (2013). Structure and activity of AbiQ, a lactococcal endoribonuclease belonging to the type III toxin-antitoxin system. *Mol. Microbiol.* *87*, 756–768.
- Sasseti, C.M., Boyd, D.H., and Rubin, E.J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* *48*, 77–84.
- Schnappinger, D., Ehrt, S., Voskuil, M.I., Liu, Y., Mangan, J.A., Monahan, I.M., Dolganov, G., Efron, B., Butcher, P.D., Nathan, C., et al. (2003). Transcriptional Adaptation of *Mycobacterium tuberculosis* within Macrophages: Insights into the Phagosomal Environment. *J. Exp. Med.* *198*, 693–704.
- Schureck, M.A., Maehigashi, T., Miles, S.J., Marquez, J., Cho, S.E., Erdman, R., and Dunham, C.M. (2014). Structure of the *Proteus vulgaris* HigB-(HigA)₂-HigB toxin-antitoxin complex. *J. Biol. Chem.* *289*, 1060–1070.
- Sheldrick, G.M., and IUCr (2008). A short history of *SHELX*. *Acta Crystallogr. Sect. A Found. Crystallogr.* *64*, 112–122.
- Slayden, R.A., Dawson, C.C., and Cummings, J.E. (2018). Toxin–antitoxin systems and regulatory mechanisms in *Mycobacterium tuberculosis*. *Pathog. Dis.* *76*.
- Smale, S.T. (2010). Beta-galactosidase assay. *Cold Spring Harb. Protoc.* *2010*, pdb.prot5423.
- Smith, J.A., and Magnuson, R.D. (2004). Modular organization of the Phd repressor/antitoxin protein. *J. Bacteriol.* *186*, 2692–2698.
- Stefan, M.I., and Le Novère, N. (2013). Cooperative binding. *PLoS Comput. Biol.* *9*, e1003106.
- Sulis, G., Roggi, A., Matteelli, A., and Raviglione, M.C. (2014). Tuberculosis: epidemiology and control. *Mediterr. J. Hematol. Infect. Dis.* *6*, e2014070.
- Tan, Q., Awano, N., and Inouye, M. (2011). YeeV is an *Escherichia coli* toxin that inhibits cell division by targeting the cytoskeleton proteins, FtsZ and MreB. *Mol. Microbiol.* *79*, 109–118.
- Thisted, T., and Gerdes, K. (1992). Mechanism of post-segregational killing by the *hok/sok* system of plasmid R1: *Sok* antisense RNA regulates *hok* gene expression indirectly through the overlapping *mok* gene. *J. Mol. Biol.* *223*, 41–54.
- Torrey, H.L., Keren, I., Via, L.E., Lee, J.S., and Lewis, K. (2016). High Persister Mutants in *Mycobacterium tuberculosis*. *PLoS One* *11*, e0155127.

- Walden, H. (2010). Selenium incorporation using recombinant techniques. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 352–357.
- Walling, L.R., and Butler, J.S. (2016). Structural Determinants for Antitoxin Identity and Insulation of Cross Talk between Homologous Toxin-Antitoxin Systems. *J. Bacteriol.* **198**, 3287–3295.
- Wang, X., and Wood, T.K. (2011). Toxin-Antitoxin Systems Influence Biofilm and Persister Cell Formation and the General Stress Response. *Appl. Environ. Microbiol.* **77**, 5577–5583.
- Wang, X., Lord, D.M., Cheng, H.-Y., Osbourne, D.O., Hong, S.H., Sanchez-Torres, V., Quiroga, C., Zheng, K., Herrmann, T., Peti, W., et al. (2012). A new type V toxin-antitoxin system where mRNA for toxin GhoT is cleaved by antitoxin GhoS. *Nat. Chem. Biol.* **8**, 855–861.
- Wayne, L.G. (1994). Dormancy of *Mycobacterium tuberculosis* and latency of disease. *Eur. J. Clin. Microbiol. Infect. Dis.* **13**, 908–914.
- Wen, Y., Behiels, E., and Devreese, B. (2014). Toxin-Antitoxin systems: their role in persistence, biofilm formation, and pathogenicity. *Pathog. Dis.* **70**, 240–249.
- Wen, Z., Wang, P., Sun, C., Guo, Y., and Wang, X. (2017). Interaction of Type IV Toxin/Antitoxin Systems in Cryptic Prophages of *Escherichia coli* K-12. *Toxins (Basel)*. **9**, 77.
- Wessner, F., Lacoux, C., Goeders, N., Fouquier d'Hérouel, A., Matos, R., Serror, P., Van Melder, L., and Repoila, F. (2015). Regulatory crosstalk between type I and type II toxin-antitoxin systems in the human pathogen *Enterococcus faecalis*. *RNA Biol.* **12**, 1099–1108.
- Winter, G., and IUCr (2010). *xia2* : an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190.
- Zielenkiewicz, U., and Ceglowski, P. (2005). The toxin-antitoxin system of the streptococcal plasmid pSM19035. *J. Bacteriol.* **187**, 6094–6105.